Inter-individual similarities in internal models of the world shape similarities in the perception and neural processing of scenes

Micha Engeser [1,2,*] & Daniel Kaiser [1,2,3,4]

[1] Neural Computation Group, Department of Mathematics and Computer Science, Physics, Geography, Justus Liebig University Giessen, 35392 Giessen, Germany

[2] Center for Mind, Brain and Behavior (CMBB), Philipps University Marburg, Justus Liebig University Giessen, and Technical University Darmstadt, 35032 Marburg, Germany

[3] Center for Applied Computer Science and Data Science (ZAD), Justus Liebig University Giessen, 35392 Giessen, Germany

[4] Cluster of Excellence "The Adaptive Mind", Justus Liebig University Giessen, Philipps University Marburg, and Technical University Darmstadt, 35392 Giessen, Germany

** Correspondence to: michaengeser@gmail.com*

Abstract

Every individual person perceives the visual world in their own unique way, yet we still know little about the origins of these individual differences. Here, we propose that idiosyncrasies in internal models—mental representations of what the world should look like—shape how individuals perceive and process natural scenes. To characterize these internal models, participants drew what they considered the most typical version of a given scene category. Using a combination of deep learning tools, we quantified inter-subject similarities in these drawings and used them to predict inter-subject similarities in perceptual task performances and neural responses to a fully independent set of natural scenes. Individuals with more similar internal models showed more similar scene categorization performance and judged several scene properties (typicality, usability, and complexity) more similarly. Moreover, variations in participants' internal models predicted inter-subject correlations in BOLD time courses in the lateral occipital and lateral prefrontal cortices. These results provide a novel, mechanistic explanation of how perceptual and neural alignment across individuals is shaped by idiosyncrasies in prior experience.

Introduction

Most research in vision science has focused on population-level effects, aiming to uncover perceptual mechanisms that are shared across individuals. Hence, variability between participants was largely treated as unpredictable noise[1]. Yet, recent studies highlight a range of systematic inter-individual differences in natural vision, spanning face perception[2,3], object recognition[4], and scene exploration[5–7]. On the neural level, individual differences in object and scene representations are found in ventral visual cortex[8,9] and the lateral prefrontal cortex (LPFC)[10,11], suggesting that both perceptual and cognitive brain systems contribute to individual perception.

Although these studies demonstrate meaningful inter-individual differences, they do not readily tell us *why* people differ. Here, we provide an explanation based on inferential theories of visual perception[12–14], which posit that the efficiency of perception hinges on the convergence of sensory inputs with internal models (or priors) of what the world should look like. On this view, similarities in perception and neural representation may be predicted by similarities in people's internal models of the world, that is, their individually specific expectations about the structure of the world.

Yet, how can we examine similarities and differences in people's internal models? Here, we use drawings as a tool to characterize internal models in a detailed and relatively unconstrained way[15–17]. By asking people to produce typical drawings of indoor scene categories, we can obtain pictorial descriptions of their internal models of that scene category. On the group level, such scene drawings capture statistical regularities of real-world scenes[18], rendering them suitable to capture such regularities on the individual-participant level as well. Further, previous work suggests that typical scene drawings indeed capture key properties of internal models: When 3D-rendered scenes are hand-crafted to be highly similar to individual participants' drawings, these scenes are better categorized than scenes that are similar to other people's drawings[19,20].

Here, we investigated whether inter-individual similarities and differences in typical scene drawings (and thus participants' internal models) predict similarities and differences in perception and neural responses, for a completely independent set of real-world scene photographs. We leveraged deep learning tools for objectively quantifying inter-individual similarities in these drawings to assess similarities in participants' individual internal models. Following an inter-subject representational similarity analysis (IS-RSA) approach[21–23], we arranged the pairwise dissimilarities in internal models into an inter-subject representational dissimilarity matrix (IS-RDM), which we in turn compared to IS-RDMs derived from behavioral performances and functional magnetic resonance imaging (fMRI) measurements.

Our study yields two critical findings: First, we demonstrate that inter-individual similarities in internal models predict inter-individual similarities in behavioral categorization performance and property ratings for a novel set of real-world scenes. Second, we demonstrated that inter-individual similarities in internal models also predict the similarity of fMRI response dynamics to these novel scenes in object-selective visual cortex and across lateral regions in frontal cortex. Together, these results demonstrate that individual priors about the structure of the world provide a parsimonious explanation for idiosyncrasies in brain and behavior.

Results

In our experiments, we characterized internal models in a drawing task[17,19,20], where participants drew the most typical instances of two scene categories (e.g., kitchens and bathrooms). We instructed participants to draw the most typical version of a scene category, but not their most preferred scene or their own living environment. Participants had 30 seconds to think about their drawing and four minutes to execute it. A predefined perspective grid was supplied to standardize viewpoints across drawings. To objectively quantify inter-individual similarities in participants' drawings (and thereby, their internal models), we computed pairwise similarities between all drawings from deep neural network (DNN) activations, using the penultimate layer of VGG16[24] trained on ImageNet[25] (Figure 1A). Given that this DNN was trained on natural images rather than drawings, we transformed all the drawings into naturalistic, pictorial versions using a diffusion-based AI tool (Draw3D; https://draw3d.online, see Figure S2 for example drawings and generated images) before supplying them to the DNN. These processing choices were validated in a supplementary experiment: DNN activations to these converted drawings, evaluated in late layers of VGG-16 trained on ImageNet, best mirrored human ratings of pairwise dissimilarity in a subset of the drawings (see Figure S1). We stored the resulting pairwise similarities in participants' typical scene drawings in an internal model IS-RDM, representing the structure of individual differences in participants' internal models.

To unequivocally attribute any of the following results to differences in internal models, our experiments also included two control conditions. First, in the copy control condition, all participants copied the same scene photographs, allowing us to estimate inter-subject variations in drawing ability and style independently of variations in internal models of a scene. In a supplementary experiment, we show that similarities in the control drawings can predict similarities in drawing style and quality ratings for the typical drawings of the same participant (Figure S1). The copy control condition also accounts for familiarity acquired during the drawing task[19], as participants familiarized themselves with both the typical and copy

drawings. Second, in the photo control condition, we asked participants to provide photographs of their private rooms (e.g., their own kitchens or bathrooms), allowing us to estimate inter-subject variations in the structure of their current living environments. This also addresses concerns that participants' immediate familiarity with a scene category, rather than their internal models, drives individual differences in brain and behavior. For both control conditions, we created separate IS-RDMs following the same DNN-based procedure as for the typical scene drawings (Figure 1A), and then partialed out the control IS-RDMs when computing correlations between the typical drawing IS-RDMs and behavioral or neural IS-RDMs.

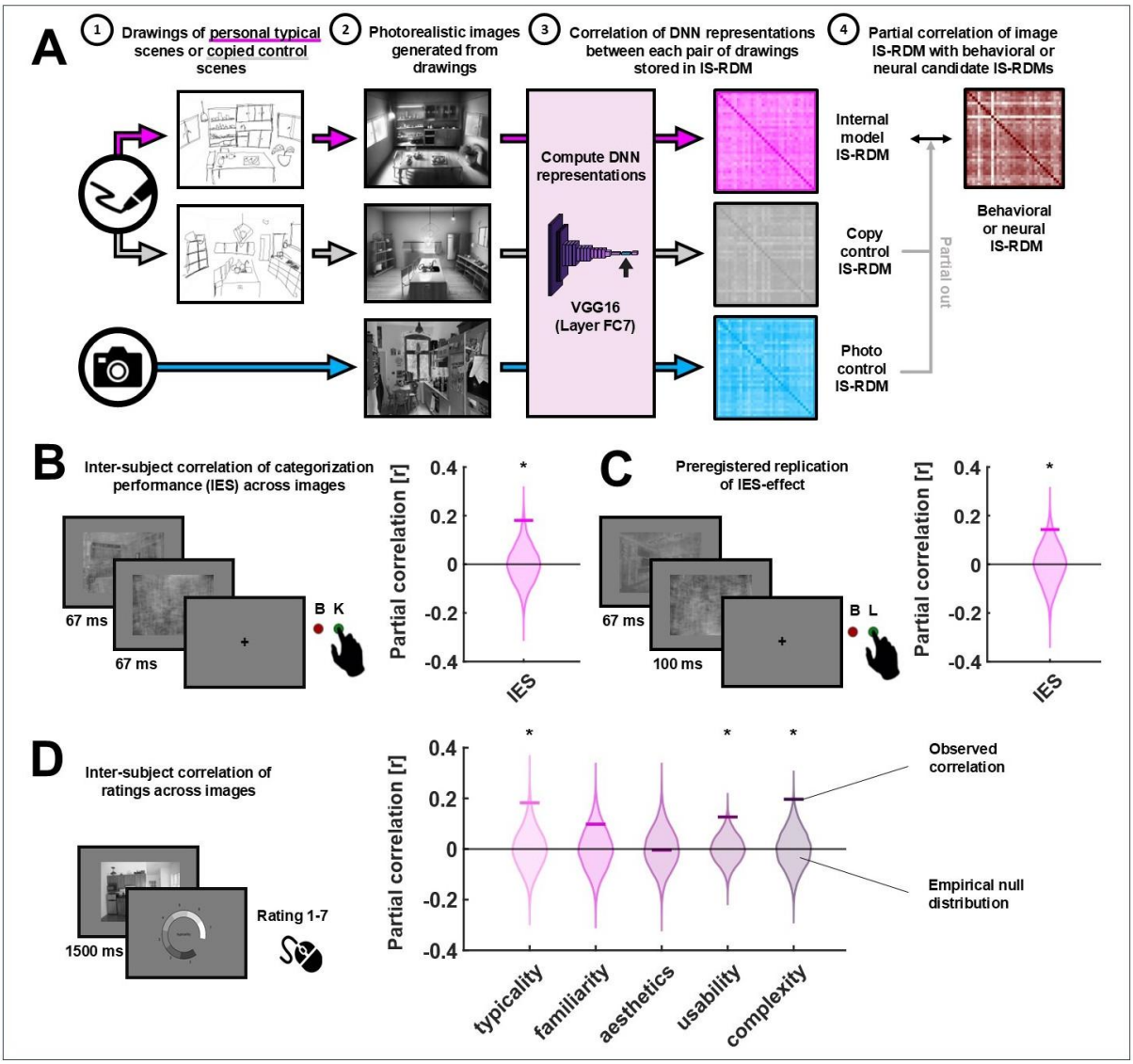

Figure 1 – Inter-subject RSA and behavioral results. (A) *Illustration of inter-subject representational similarity analysis (IS-RSA). First, participants drew their most typical version of a scene category serving as descriptors of their internal models (pink). Next, we generated photorealistic images from these drawings using a generative AI model. To measure inter-subject similarities in these images, we used pairwise correlations of late-layer feature activations in a VGG16 DNN. The resulting inter-subject representational dissimilarity matrix (IS-RDM) was then correlated with candidate IS-RDMs to test whether inter-subject similarities in internal models can predict how the same individuals differ in behavioral or neural metrics.*

We used partial correlations throughout, controlling for individual differences in drawing ability and style (grey) and recent visual experience (blue). In the copy control condition, all participants copied the same scene photograph. An IS-RDM of the copy drawings was constructed in the same way as for the typical drawings, representing how these individuals differed in their drawing abilities and style. In the photo control condition, we correlated VGG16 representations of photos of participants' private rooms. ***(B)*** *Results of the categorization experiment. Participants categorized briefly presented and masked images of bathroom and kitchen scenes. To quantify inter-subject similarities, we correlated how individuals' categorization performances (IES, inverse efficiency score) differed across the stimulus set. Next, we correlated the resulting IS-RDM with the internal model IS-RDM, partialing out the control predictor IS-RDMs.* ***(C)*** *Results of the preregistered replication of the categorization experiment, instead using bedroom and living room scenes. Across both experiments, inter-individual differences in internal models predicted inter-individual differences in categorization performance.* ***(D)*** *Results of the scene property rating experiment. Here, participants rated five properties for every scene on a 1-to-7 scale. Inter-individual differences in internal models predicted inter-individual differences in typicality, usability, and complexity ratings. Violin plots show the empirical null distributions obtained from permutation tests, and the horizontal bars show the observed correlation coefficients. Asterisks indicate p<0.05 (permutation test, FDR-corrected for multiple comparisons where applicable).*

Similarities in internal models predict similarities in scene categorization

First, we tested whether inter-individual similarities in internal models predict inter-individual similarities in behavioral categorization performance. Previous studies provided the first evidence for such a relationship, showing that 3D scene renders hand-crafted to mirror individual participants' drawings are more accurately categorized[19,20]. Here, we go a step further and ask whether we can use typical scene drawings to predict categorization performance for a completely different, novel set of natural scene images.

All participants (N = 35) first completed a drawing session, where they drew typical kitchens and bathrooms, copied kitchen and bathroom photographs, and provided pictures of their private kitchens and bathrooms. They then participated in a challenging scene categorization task. They viewed briefly presented (67ms) and backward-masked low-contrast photographs of kitchens and bathrooms (matched for luminance and spatial frequency) and had to indicate which of the two categories they had just seen. Given the unreliable sensory information under these viewing conditions, perception should rely strongly on predictions from internal models[26,27].

Participants attained a mean categorization accuracy of 68.0 % (± 7.1%) and a mean response time of 853 ms (± 145 ms). Combining accuracy and response times, we calculated the inverse efficiency score (IES) and built the IS-RDM by correlating IES across all images of the same category between every pair of subjects (i.e., testing to what extent the same images were hard or difficult for each pair of participants). As hypothesized, this categorization IS-RDM was positively correlated with the internal model IS-RDM (Figure 1B, $r = 0.18$, $p = 0.021$, permutation test). This correlation emerged despite partialling out both control IS-RDMs, and

no similar correlations were found when assessing the copy and photo control conditions separately (Figure S4).

To assess the robustness of these results, we conducted a preregistered replication experiment (N = 40). In this experiment, we further probed the generalizability of our findings, using bedroom and living room scenes instead of the kitchen and bathroom scenes. All other aspects of this experiment were identical to the previous experiment, with the exception that the stimuli were presented for 100 ms, given that bedrooms and living rooms were harder to categorize. Participants attained a mean categorization accuracy of 71.9 % (± 7.2%) and a mean response time of 825 ms (± 123 ms). Critically, we replicated the correlation between the categorization IS-RDM and the internal model IS-RDM (Figure 1C, r = 0.14, p = 0.035, permutation test). The effects in both experiments were also confirmed in a linear mixed-effects modelling analysis, emphasizing statistical robustness (see model outputs in Table S1). Together, our categorization experiments demonstrate that inter-individual similarities in scene categorization are shaped by inter-individual similarities in internal models for these scenes. Our preregistered replication further suggests that this relationship generalizes across different scene categories.

Similarities in internal models predict similarities in scene property ratings

Next, we investigated whether inter-individual similarities in internal models could also predict behavior beyond a (deliberately) challenging categorization task regime. After our first categorization experiment, participants therefore also viewed the original high-quality scene photographs for 1,500 ms and rated them on a range of scene properties: typicality, familiarity, aesthetic attractiveness, usability, and complexity. The properties were selected to reflect varying subjectivity and abstractness. For each of these properties, we again constructed an IS-RDM by correlating the ratings across all images for each pair of participants, which was in turn compared to the IS-RDM of internal models (while again controlling for the copy and photo IS-RDMs). Relationships among the different ratings, as well as among the resulting IS-RDMs and the categorization IS-RDM, can be found in Figure S3.

We first assessed a relationship with typicality to test that our drawing method was able to capture inter-subject variation in perceived typicality. This was expected given that participants were specifically instructed to draw the scene that was most typical for them. Indeed, the internal model IS-RDM was significantly correlated with the typicality rating IS-RDM (Figure 1D, r = 0.18, p = 0.021, permutation test, FDR-corrected). By contrast, the familiarity ratings IS-RDM yielded no significant correlation (r = 0.10, p = 0.140, permutation test, FDR-corrected), suggesting that similarity in internal models predisposes similar typicality judgments independently of familiarity effects.

We also observed a significant correlation between the internal model IS-RDM and the usability rating IS-RDM (r = 0.13, p = 0.021, permutation test, FDR-corrected) and the complexity rating IS-RDM (r = 0.20, p = 0.019, permutation test, FDR-corrected), suggesting that individual differences in how people judge scene functionality[28,29] and complexity[30,31] can be traced back to individual differences in internal models. In contrast, despite substantial inter-individual variation in aesthetic preferences for natural scenes[32,33], we found no

correlation between the internal model IS-RDM and the aesthetic attractiveness IS-RDM ($r < 0.01$, $p = 0.512$, permutation test, FDR-corrected).

To summarize our behavioral experiments, we found that individuals with more similar internal models exhibit significant commonalities in how they categorize scenes under visually challenging conditions, as well as in how they evaluate the typicality, usability, and complexity of these scenes.

Similarities in internal models predict similarities in neural dynamics across visual and frontal cortices

To test how individual internal models shape scene processing on the neural level, we next related inter-individual similarities in internal models to inter-individual similarities in cortical responses during scene viewing. In an fMRI experiment, participants ($n = 30$) viewed a series of kitchen and bathroom photographs. In 15% of trials (target trials), they were asked to indicate whether a visually presented word cue corresponded to an object featured in the previous scene image. Participants answered correctly on 63.2% (± 7.2%) of target trials, rendering this a demanding task that required paying close attention to the scenes. Importantly, the trial sequence, including the position of the target trials, was identical across participants, yielding comparable neural time series across participants.

We assessed inter-individual similarities in these time series in an inter-subject correlation (ISC) analysis (Figure 2A). In ISC analysis, neural response time courses to a time-locked stimulation are averaged within a particular ROI and subsequently correlated across individuals, thereby isolating stimulus-induced responses in individual participants from unrelated noise[34]. The pairwise ISCs form an IS-RDM, representing individual differences in how neural responses in a given brain area fluctuate with the stimulus time series. If the two participants process a certain scene differently, their neural activation would diverge at that given time point, and the ISC would be lower between these participants, yielding a higher dissimilarity in the IS-RDM. By correlating the resulting neural IS-RDMs with the internal model IS-RDMs, we delineate how much these individual processing dynamics in a brain region can be predicted by idiosyncratic internal models. As in the behavioral experiments, we again partialed out the copy and photo control IS-RDMs.

In a region-of-interest (ROI) analysis, we focused on regions in the ventral visual cortex and the prefrontal cortex, where previous studies indicated individually unique object or scene representations[8–11]. We extracted neural IS-RDMs from five anatomically and functionally defined ROIs: early visual cortex (V1), object-selective lateral occipital cortex (LOC), scene-selective parahippocampal and occipital place areas (PPA and OPA), and lateral prefrontal cortex (LPFC). ISCs were reliable in all ROIs, with significant positive correlations across halves of the experiment (see Figure S5).

We found significant correlations between the internal models IS-RDM and the neural IS-RDMs in object-selective LOC (Figure 2B, $r = 0.14$, $p = 0.018$, permutation test, FDR-corrected) and LPFC ($r = 0.13$, $p = 0.013$, permutation test, FDR-corrected) but not in V1 ($r = 0.08$, $p = 0.112$, permutation test, FDR-corrected), PPA ($r = 0.08$, $p = 0.136$, permutation test, FDR-corrected), or OPA ($r = 0.06$, $p = 0.213$, permutation test, FDR-corrected). These results indicate that individual differences in internal models shape idiosyncratic neural dynamics at

higher-level stages of visual processing and in frontal circuits. In the visual cortex, these effects emerged in object- but not scene-selective areas, suggesting that internal models could exert their influence at the level of objects and their configurations. Scene-selective regions also showed no significant correlations when similarities in internal models were computed using a scene-trained VGG16 (see Fig. S5B).

Beyond the pre-specified ROIs, we also performed a searchlight analysis across a set of 180 brain parcels[35]. Our searchlight analysis revealed 13 significant parcels (Figure 2C). Most of them were located within the LPFC (see Table S2 for exact parcel labels, correlation coefficients, and p-values). Specifically, we found significant correlations in the dorsolateral prefrontal cortex (dlPFC), inferior frontal cortex (iFC), and orbitofrontal cortex (oFC). Other significant parcels included the anterior and posterior cingulate and insular cortex. In the ventral visual stream, the only significant parcel overlapped with the LOC ROI. Our whole-brain analysis thereby confirmed that individual differences in internal models shape processing in object-selective cortex while providing a more nuanced mapping of the effects within LPFC.

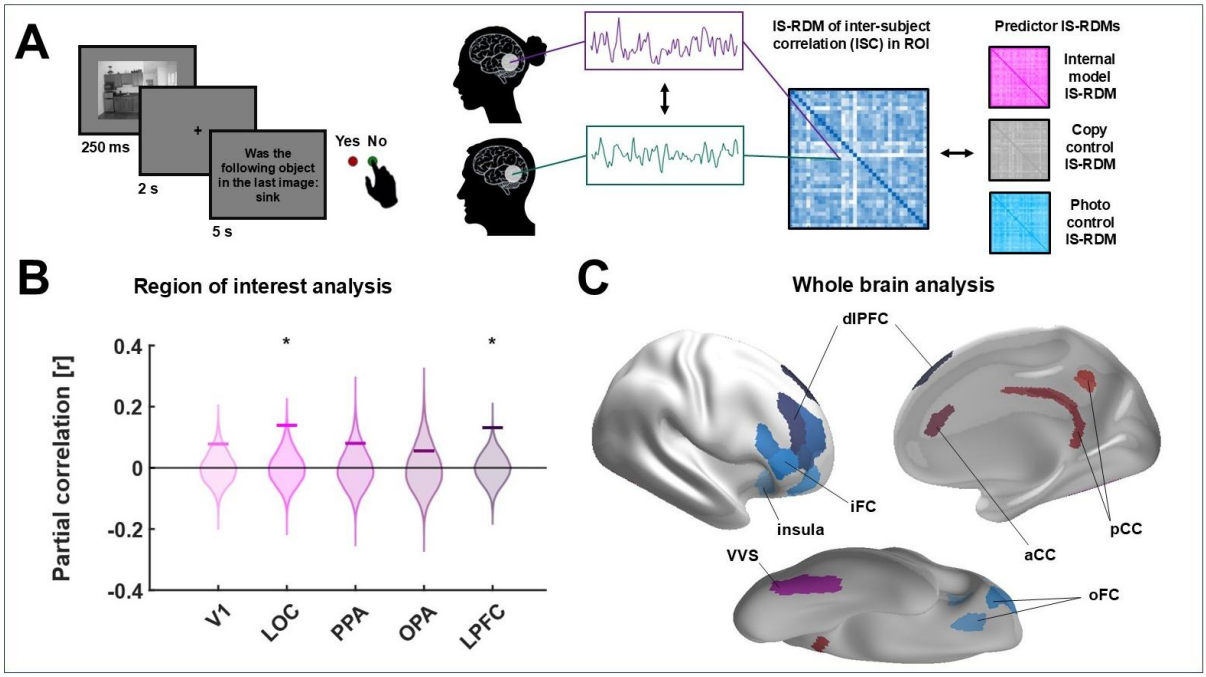

Figure 2 – fMRI results. (A) *Participants viewed a sequence of scene photographs in the MRI scanner. In 15% of trials, they were asked whether a cued object was present in the previous scene image. The recorded BOLD activity was correlated between each pair of participants, resulting in IS-RDMs reflecting similarities in brain dynamics. Neural IS-RDMs were correlated to the IS-RDM of internal models (pink), partialing out the copy (grey) and photo (blue) control IS-RDMs.* ***(B)*** *ROI results. Inter-individual differences in internal models predicted idiosyncratic brain dynamics in object-selective LOC and LPFC, but not in V1 and scene-selective PPA/OPA. Violin plots show the empirical null distributions obtained from permutation tests, and the horizontal bars show the observed correlation coefficients. Asterisks indicate p<0.05 (permutation test, FDR-corrected for multiple comparisons).* ***(C)*** *Whole-brain results. Here, the analysis was conducted across a set of 180 bilateral brain*

parcels. Only significant parcels (p < 0.05, permutation test, FDR-corrected) are displayed, including dorso-lateral prefrontal cortex (dlPFC), inferior frontal cortex (iFC), orbitofrontal cortex (oFC), insula cortex, anterior and posterior cingulate cortex (aCC and pCC), as well as ventral visual stream (VVS).

Discussion

Here, we demonstrate that inter-individual differences in the perception and processing of natural scenes are shaped by inter-individually different internal models. We characterized internal models at the individual level using a drawing task that probed participants to draw the most typical versions of natural scene categories. By combining AI-based tools for evaluating the similarity of these drawings across participants with control conditions that excluded confounds in drawing ability and scene familiarity, we devised comprehensive, image-computable descriptors of how internal models differ across people. Leveraging an IS-RSA approach, we found that individual differences in typical scene drawings can predict idiosyncrasies in the performance during categorization of an independent set of scene photographs. A preregistered replication revealed the same effect for two different scene categories. In a scene property rating task, inter-subject similarities in how participants judge the typicality, usability, and complexity (but not familiarity and aesthetic attractiveness) of scenes demonstrated a similar link to internal models. Moreover, inter-individual similarities in internal models could predict idiosyncrasies in fMRI responses in object-selective LOC and LPFC.

On the group level, it is well established that the visual system is more efficient in processing objects and scenes when they are predictably structured[36–40]. Predictive processing frameworks explain the efficiency in processing structured scenes with the probabilistic integration of sensory inputs with strong prior expectations of the world, governed by internal models learned throughout our lifetimes[26,27,41]. However, the nature of these internal models may differ between individuals, as a function of their visual diet[1,42,43] or the architecture of their visual system[44,45]. Our study provides evidence that such differences matter: When internal models of the world differ, perception and neural representation differ alongside, in systematic ways.

In our behavioral experiments, we found that people with similar internal models showed similar performance patterns in a taxing scene categorization task. This finding is consistent with previous research that shows a facilitation of categorization performance for 3D-rendered scenes hand-crafted to be similar to participants' drawings[19,20]. In contrast to this previous work, we used inter-individual similarities in the drawings to predict inter-individual similarities in categorization for a completely novel set of natural scene photographs. This establishes an exciting method that predicts similarities and differences in perceptual performance across individuals using a single drawing from each participant, which can be obtained in a few minutes. Future studies may routinely acquire such drawings to predict individual differences in perceptual, cognitive, and action-related tasks featuring objects or scenes.

We further show that the predictive power of individual differences in internal models, as measured by drawings, exceeds challenging perceptual tasks like the categorization of degraded stimuli. In a scene property rating task, we found that judgments of typicality, usability, and visual complexity are more similar among individuals with more similar internal

models. This shows that internal models provide rich information that transcends features relevant for categorization. Interestingly, judgments of scene familiarity were not influenced by individual differences in internal models. This result corresponds with our control conditions, which probed short-term familiarity acquired during drawing (copy control) and familiarity with participants' lived environments (photo control), and which failed to predict categorization accuracy and most of the rating data (see Figure S4). Typicality and familiarity can therefore be disentangled, and our drawing-based task is capable of isolating the contribution of typicality.

On the neural level, we show that inter-individual similarities in fMRI response dynamics to a novel set of scene images are predicted by similarities in internal models. In the visual system, the only region that showed a significant correlation of ISC and individual differences in internal models was object-selective LOC, which was corroborated in the whole-brain analysis. No such effect emerged in V1, in line with the idea that semantic or contextual expectations about natural environments prominently exert their influence on later stages of the processing hierarchy[26,27,46]. Response dynamics in scene-selective OPA and PPA were not influenced by idiosyncrasies in internal models either, even when evaluating drawing similarities with scene-trained DNN models. The finding that object-selective but not scene-selective regions showed a significant effect suggests that individual differences in internal models most strongly hinge on how individuals differently weigh individual objects and their spatial relations[47–50]. One reason for this dominance of object information could be that we enforced a fixed perspective during drawing, somewhat eliminating variations in the global spatial configuration of rooms. Scene-selective regions may be specifically sensitive to such global configurations[51,52] and how they differ across individuals. Another reason could be that our task necessitated the encoding and memorization of object identities, which may have prioritized local object information over global scene information.

Interestingly, individual differences in internal models also predicted idiosyncratic response dynamics outside of the visual cortex. Our ROI analyses revealed significant effects in the LPFC. These effects were corroborated by the whole-brain analysis, which additionally revealed significant effects in the cingulate cortex. Particularly, LPFC shows large inter-subject variability in terms of anatomical structure[44,53], functional connectivity[53,54], as well as the representation of visual content[10,11] and abstract concepts[55]. Most relevantly, Lin and Lau[11] found that the representation of natural scenes is more individualistic than in the visual cortex and hypothesized that these individual differences are the result of random projections between sensory and high-level brain regions to facilitate flexible working memory. Here, we demonstrate that these individual differences in LPFC activity are systematically related to the subject's personal scene priors. While connections from the visual cortex to LPFC might indeed be an interface to visual memory (consistent with the dlPFC identified in our searchlight analysis)[56], this interface may also operate in the top-down direction, where prior knowledge is activated and routed to the visual cortex in the form of predictions. This idea is consistent with earlier proposals, which suggest that LPFC predictively biases object processing in ventral visual cortex by a rapid top-down propagation of experience-based priors[41,57,58]. This idea resonates well with the current findings, suggesting that an interplay of LPFC and LOC mediates the influence of internal models on scene processing.

Yet, the necessity of frontal involvement for maintaining adaptive perception is hotly debated[59–63]. On the one hand, converging evidence suggests that the PFC primarily plays a role under

ambiguous or demanding stimulus and task conditions[64–69]. The object-related task our participants performed in the MRI was indeed judged to be very challenging in an exit questionnaire, and most probably required high attentional and working memory capacity. Thus, our LPFC results could potentially reflect individual differences in how participants attribute attention and engage working memory across the presented stimuli as a function of their personal prior knowledge. For example, if a scene is particularly unpredictable for an individual, their frontal regions may be involved more prominently compared to an individual whose priors align better with the respective scene.

On the other hand, the dorsolateral and inferior frontal regions identified in our whole-brain analysis have also been repeatedly associated with conscious access, partially even in the absence of overt task responses[61,70–72]. Hence, the observed LPFC patterns may reflect individual differences in conscious perception. Recent work by Kawakita and colleagues[73] proposed that comparing the similarity structures of subjective experience across individuals can provide an empirical measure of how phenomenal properties align between observers. From this perspective, our findings may also suggest that individual differences in internal models manifest in the qualitative structure of visual experience and their neural correlates. These insights open promising avenues for investigating atypical perception in clinical populations—such as psychosis—where internal models may deviate systematically from those of the neurotypical population and thereby give rise to atypical conscious experiences of the world.

Taken together, our findings show that individual differences in internal models predict inter-subject variability in categorization performance, judgments of scene properties, and neural processing dynamics in both visual and prefrontal cortex. These results complement prior work linking perceptual idiosyncrasies to biological “hardware”—from genes to cortical anatomy. We demonstrate that also individual differences in visual “software”, that is, how the brain computes inferences about the world, play a critical role in shaping idiosyncrasies in perception. Our study therefore provides a mechanistic account of not only *how* but also *why* perception and its neural underpinnings systematically differ between individuals.

Methods

Experiment 1

Participants

In the first experiment, we recruited 41 participants (21 female, mean age 28.4 ± 7.0). Six were excluded from the analysis: three did not complete the second session, two did not achieve significant above-chance level categorization accuracy, and one was excluded due to insufficient drawing quality (see Figure S2B for excluded drawings). This resulted in a sample size of 35 participants, equivalent to previous studies that used drawings to capture differences in internal models[19]. All participants reported normal or corrected-to-normal vision. Participants were recruited through the participant recruitment system of the Justus Liebig University Gießen, provided written informed consent, and received monetary reimbursement

of ten Euros per hour. Procedures were approved by the General Ethical Committee of the Justus Liebig University Gießen and adhered to the Declaration of Helsinki.

Procedure

The experiment consisted of two sessions. The first session contained the drawing task (~ 30 min. duration) and the categorization task (~ 90 min.), and the second session the rating task (~60 min.) and an exit questionnaire (~15 min.). Before the experiment, participants provided photographs of their own bathrooms and kitchens. Photos should be in landscape format and capture as much of the room as possible. In cases where the entire room could not be captured in a single image (for example, when the toilet and shower were located in separate rooms), participants could provide more than one photo per category. All experiments were controlled using Psychtoolbox-3[74] on MATLAB, if not stated otherwise.

Drawing Task

To characterize internal models, we employed the drawing task from Wang and colleagues[19]. In brief, participants drew what they thought was the most typical version of an indoor scene for two categories (bathroom and kitchen). Participants had 30 seconds to imagine their typical scenes and four minutes to draw them. All drawings started with a template showing the outline of a room in which participants drew the imagined scene, ensuring comparable viewpoints across people. In a practice round, participants first drew a typical classroom to familiarize themselves with the task and drawing equipment. Additionally, we included a copy control condition in which participants copied a scene photograph of the two categories under the same time restriction. Individual differences in the control drawings served as an estimate of inter-subject variance in drawing-related confounds, such as drawing ability and style, and were partialed out from the individual differences in the typical drawings. All drawings were made on an Apple iPad Pro (19.5 cm by 26.1 cm display) using the Sketchbook app and an Apple Pencil. Participants could use the standard pen option to produce black lines (fixed width) and the eraser to delete lines.

Categorization task

After the drawing task, all subjects performed a categorization task in which they had to categorize briefly presented photographs of bathrooms and kitchens. Fifty images for each category were selected from the SUN Database[75], spatial frequency and luminance matched using the default settings of the SHINE toolbox[76], and presented at a visual angle of 8° × 5°, and a viewing distance of 55 cm. Each trial began with a fixation period, during which a fixation cross was displayed for a jittered duration (averaging 1 second, sampled from an exponential distribution truncated at 0.7 and 2 seconds). Next, an image was presented on the screen for 67 ms, followed by a mask of 67 ms duration consisting of an average image across all stimuli in the experiment. Participants had to indicate which scene category had been presented by pressing one of two response keys on a keyboard. The fixation cross was shown while waiting for the response. Each image was repeated 12 times, yielding a total of 1,200 trials, including self-paced breaks after every 50 trials. The stimulus order was pseudo-randomized once and fixed across participants.

Rating task

In the rating task, participants rated the same images used in the categorization task (without spatial frequency and luminance normalization) based on five different criteria (typicality, familiarity, aesthetic attractiveness, usability, and complexity) on a scale of 1 to 7. First, all bathroom images were rated, followed by all kitchen images. Ratings of the five rating criteria were provided in separate blocks (in the order listed above), in which stimuli were shown for 1500 ms, followed by a response wheel displaying the response options in clockwise order with a random starting point[77]. Responses were provided by clicking on the respective number with the mouse cursor. Participants had to rate the images relative to each other within one category; therefore, all images of the respective category were shown on the screen in two batches of 25 images for a self-paced duration before the ratings started. The rating criteria were defined as follows: (1) typicality - “Provide your personal judgment on the extent to which the presented scene is a typical version of its scene category (bathroom/kitchen)”, (2) familiarity - “Provide your personal judgment on the extent to which the presented scene resembles places you have lived in or where you spent a substantial amount of time during your life”, (3) aesthetic attractiveness - “Provide your personal judgment on the aesthetic appeal of the scene. In other words how beautiful the scene looks to you.”, (4) usability - “Provide your personal judgment on the usability/practicability of the presented scene if you were to interact with it”, (5) complexity - “Provide your personal judgment on the complexity of the presented scene. You can think of complexity as the number of different elements appearing in a scene”.

Analysis

In all experiments, we employed an inter-subject representational similarity analysis (IS-RSA) framework[54] to investigate whether similarities in internal models of the world predict individual variations in scene perception. Inter-individual differences were captured in an inter-subject representational dissimilarity matrix (IS-RDM) of size n x n (where n is the sample size of the respective experiment), quantifying the inter-subject variation for each variable of interest, such as drawings, categorization performance, and ratings. Separate IS-RDMs were generated for the two scene categories used.

Inter-subject similarities in drawings

Relating individual differences in internal models (captured through drawings) to individual differences in perception requires estimating the similarity between the drawings of individuals. Here, we objectively estimated these similarities by comparing feature activations of a deep neural network (DNN) as a proxy measure for human perceptual similarity. As we used DNNs pre-trained on photographs, we first generated photorealistic images from the drawings using a diffusion-based AI method, the commercial Draw3D software. Each drawing was uploaded to the web application, and a prompt was added in the following format: "Produce a highly photorealistic, high-quality image of a [scene category], incorporating the following objects: [list of all objects in the drawing from left to right]. Pay special attention to realistic lighting, textures, and spatial arrangement to achieve a lifelike representation." Representative examples of images generated from the drawings are shown in Figure S2.

For each drawing, the image generation was repeated until three images of satisfactory quality were produced. Quality was judged by the experimenter, ensuring that each object in the drawings was accurately depicted in at least one of the three generated images and that the images had a photorealistic appearance. Image generation was completed before inspecting any of the subsequent analysis results.

The generated images were then presented to a deep neural network (VGG-16 trained on ImageNet). This model revealed the best match to human ratings of similarity in the content of the drawings during an experiment comparing a selection of different networks and training materials (see supplementary experiment, Figure S1). The feature maps extracted from the second-to-last fully connected layer “fc7” were correlated (Spearman correlation) between each pair of drawings. For each pair of subjects, the dissimilarities (1 - R) in the typical and copy control drawings were averaged across the three instances of generated images and stored in two separate IS-RDMs.

Inter-subject similarities in photos

To estimate similarities in participants' everyday experiences with bathrooms and kitchens, we also created IS-RDMs from the personal photographs participants provided based on VGG-16 layer “fc7” activations. Therefore, extracted feature maps were correlated between each pair of subjects and subtracted from one to create an IS-RDM separately for each category. When more than one photo was provided for the same category, the dissimilarities were averaged across them. These IS-RDMs were also partialled out from the correlation of typical drawings and behavioral IS-RDMs.

Inter-subject similarities in behavioral responses

The categorization and rating tasks aimed to measure individual differences in how participants perceive bathroom and kitchen scenes. The inter-subject dissimilarities in these behavioral responses were quantified by calculating 1 - the Spearman correlation of categorization performance or ratings across the 50 images of the same scene category for each pair of subjects. The logic here is that participants with a more similar perception should produce a more similar rank ordering across images (e.g., the same scenes should be easy or hard to categorize for them).

For the categorization task, we combined reaction time and accuracy data into an inverse efficiency score (IES) for each image by dividing the mean reaction time by the mean accuracy. To avoid division by zero, accuracy values of zero were replaced by 1/24 (equivalent to half the accuracy value if only one out of the 12 repetitions of that image was correctly categorized). Correlations of the resulting IS-RDMs of each task with each other task aimed to test how much inter-individual differences generalized across the different tasks (see Figure S3).

Comparison of inter-subject similarities in drawings and behavioral responses

Lastly, we tested our hypothesis that participants with more similar internal models of bathrooms and kitchens perceive photographs of these scenes in more similar ways. Therefore, we examined how well the IS-RDMs of drawings and photos could predict IS-RDMs of the different behavioral responses using partial Pearson correlations of the off-diagonal

values in the lower triangle of the IS-RDMs. The main analysis contained a partial correlation of the IS-RDMs of a behavioral metric and the IS-RDMs of typical drawings, controlling for inter-subject similarities in the copy control drawings and the photos. Supplementary Figure S4 shows the partial correlations for the control predictors partialling out other control predictors, and the typical drawings IS-RDMs. Separate correlation coefficients were calculated for the two categories and then averaged.

Statistical analysis

We employed a non-parametric permutation test with 10.000 subject-wise permutations when correlating IS-RDMs. At each iteration, the rows and columns of the reference IS-RDM were shuffled randomly while the intercorrelations between observations were kept intact. The p-value was obtained by calculating the proportion of permutations in which the correlation coefficient was equal to or higher than the one observed in the experiment. A p-value below 0.05 was considered a significant result.

Additionally, we employed a linear mixed-effect model with partially crossed random effects of subjects[78] to confirm the statistical inference from the partial correlation of IS-RDMs of categorization performance with typical drawings, control drawings, and photos. Thereby, the dependent variable was the pairwise inter-subject correlation (ISC) in categorization performance across all pairs of subjects, while inter-subject similarities in typical drawings, control drawings, and photos, as well as their interactions, were included as fixed effects. Inter-subject similarities were centered and treated as numerical values. Three random effects were included as nominal factors: the scene category and the identities of the two subjects of the respective pair, each with random intercepts and slopes for the fixed effects. Likelihood-ratio chi-square (χ^2) tests of the full model against the model without the respective effect were used to obtain p-values (see Table S1 for model output). The model can be summarized in the following formula:

$$ISC_{IES} \sim ISC_{typical\ drawings}\ x\ \ ISC_{control\ drawings}\ x\ ISC_{photos}\ +$$

$$(1\ x\ ISC_{typical\ drawings}\ x\ ISC_{control\ drawings}\ x\ ISC_{photos}\ |\ category)\ +$$

$$(1\ x\ ISC_{typical\ drawings}\ x\ ISC_{control\ drawings}\ x\ ISC_{photos}\ |\ subject\ 1)\ +$$

$$(1\ x\ ISC_{typical\ drawings}\ x\ ISC_{control\ drawings}\ x\ ISC_{photos}\ |\ subject\ 2)$$

Experiment 2

The first experiment was carried out in an exploratory manner. To confirm the findings of the comparison between inter-subject similarities in the drawings and the categorization task from the first experiment, we conducted a pre-registered replication (https://osf.io/zjtwx/) using two new indoor scene categories: bedrooms and living rooms. The methods of the second experiment mirrored those of Experiment 1, with a few changes described below.

Participants

For the second experiment, we recruited 43 participants (21 female, one diverse, mean age 26.3 ± 2.2), of whom three were excluded due to insufficient drawing quality, corresponding to the preregistered sample size of 40.

Procedure

All participants performed the same drawing task and the categorization task in a single session. The rating task and the exit questionnaire were not included in the second experiment, and participants did not provide photos of their private places.

Drawing Task

The drawing task followed the same procedure as in Experiment 1, with typical and control drawings of bedrooms and living rooms. As a practice, participants drew a kitchen.

Categorization task

For the categorization task, 50 bedroom and living room images were selected from the SUN Database and presented as described above, with the exception that the presentation time of the images was 100 ms.

Analysis

The analysis was identical to the analysis for Experiment 1. Here we declare that the preregistration had to be updated after data collection (updated and original registration can be found here: https://osf.io/zjtwx/). The originally registered parametric t-test-based approach is flawed as it assumes independent observations, while the values in IS-RDMs have strong interdependencies. This test was replaced by the permutation test as described for Experiment 1, which respects the interdependencies in the IS-RDMs and is generally regarded as more conservative than t-test-based methods[79]. Moreover, a linear mixed modelling analysis was used to confirm these results, following Chen et colleagues[78].

fMRI experiment

Participants

In the fMRI experiment, 33 participants (16 female, one diverse, mean age 27.7, ± 3.8) were recruited. Three of them did not complete the full MRI session and were excluded, leaving 30 participants for analysis. This sample size was based on a recent reliability analysis[80], which showed that inter-subject correlations at this sample size converge closely to those of larger samples. All participants were right-handed, healthy adults with normal or corrected-to-normal vision, received monetary reimbursement, and provided written informed consent. The study was approved by the General Ethical Committee of the Justus Liebig University Gießen in accordance with the Declaration of Helsinki.

Procedure

Drawing Task

All subjects performed the drawing task as described in the first experiment. The scene categories were bathroom and kitchen, and for practice, we used bedroom drawings. Participants also provided photos of their private bathrooms and kitchens.

Experimental design

During the fMRI scans, participants viewed the same grayscale, high-resolution bathroom and kitchen images (50 examples each) used in the ratings task of the first experiment. Each of the 12 runs of the experiment consisted of 100 images, yielding a total of 1200 trials and 12 repetitions of each stimulus. In each run, two repetitions of all stimuli of one category were presented alternating between bathroom and kitchen runs. A trial started with a 2s fixation period showing a black fixation cross, followed by a 250 ms stimulus presentation at 8° × 6° visual angle. Following 15% of the trials in each run, a task screen appeared asking the participant whether a specific object appeared in the trial just before the question (i.e., “Was the following object in the last image: sink”). The target was present in half the target trials in each run. The participant had a 5s response window to respond using a two-button response box and received feedback about the average accuracy at the end of each run. The task ensured that participants paid attention to the presented stimuli, as they did not know which scene would be followed by a task screen, while restricting responses to fixed time windows to avoid overlapping of motor responses with visual processing of the stimuli. Stimulus order and target trial positions were randomized once and kept fixed across participants. Target objects were handpicked.

fMRI acquisition and preprocessing

MRI data were acquired using a 3T Siemens Magnetom PRISMA scanner (Siemens, Erlangen, Germany) and a 64-channel head coil. The session began with a high-resolution (1 mm^3 voxel size) T_1-weighted anatomical scan, acquired using MPRAGE. Next, T_2*-weighted functional images were collected using a gradient-echo EPI sequence (TR = 1,850ms, TE = 30ms, 75° flip angle, 2.2 × 2.2 × 2.2 mm voxel size, 58 slices, 20% gap/distance factor, 220mm FOV, 100 × 100 matrix size, descending acquisition). We recorded a functional localizer (390 volumes, blocked design using videos of faces, objects, and scenes as well as scrambled videos, adapted from Kiniklioglu[81]) and 12 functional task runs (188 volumes each).

Preprocessing of the MRI data was conducted in SPM12 (www.fil.ion.ucl.ac.uk/spm). The functional volumes were first realigned and coregistered to the T_1 image, followed by normalisation of both the anatomical and functional images to Montreal Neurological Institute (MNI) standard space.

Analysis

Inter-subject similarities in drawings

Processing and analysis of the drawings and photos followed the first experiment.

Region of interest definitions

For a spatially targeted analysis of fMRI responses, we defined a set of regions of interest (ROIs) spanning the visual hierarchy from early visual cortex up to the prefrontal cortex. The early visual cortex (V1) region was masked anatomically based on Rosenke et al.[82]. The other visual ROIs were defined functionally, using a general linear model (GLM) implemented in SPM 12 to model the fMRI responses in the localizer run, separately for each participant. Contrasts were calculated between stimulus categories, and each participant's top 200 voxels with the highest t-values from the respective contrast were selected within a probabilistic anatomical mask[83]. For scene-selective regions (PPA and OPA), a scenes > objects contrast was used, while for the object-selective LOC, an objects > scrambled contrast was used. For the lateral prefrontal cortex (LPFC), we used the same liberal LPFC mask as Lin and Lau[11] and identified the 200 most visually responsive voxels within that mask using an intact > scrambled contrast. Figure S7A shows that the IS-RSA results of our ROI analysis are consistent across different numbers of maximally responsive voxels selected.

Inter-subject correlation of neural time courses

To assess idiosyncrasies in neural responses to the scene stimuli, we conducted an inter-subject correlation (ISC) analysis. ISC represents the Pearson correlation of the activity time course in a spatial location (ROI- or parcel-wise) across all pairs of subjects[34]. For the ROI analysis, the time courses were averaged over all voxels in the ROI, leaving out the first and last three TRs. Next, time courses were detrended (removing the best straight-line fit linear trend), and the group mean was linearly regressed out from each individual's data to boost idiosyncratic responses[21]. ISCs were computed separately for each run and averaged over the six runs of the same category.

A split-half correlation provided an estimate of the reliability of the ISC across runs. Specifically, the ISC was computed separately for the two halves of the runs and then correlated (Pearson) between the halves. The mean across all possible run-splits was calculated separately for each scene category and subsequently averaged. Statistical significance was assessed using a permutation test, in which the IS-RDM for one half was shuffled on a subject-wise basis (see above for detailed description). The resulting p-values were FDR-corrected (Figure S5).

To test the relationship between individual differences in neural response to scene stimuli and variability in internal models, the IS-RDMs of each ROI and voxel were compared to the predictor IS-RDMs of typical drawings, partialling out control drawings, as well as the provided photos. Partial correlations with the control predictors are shown in Supplementary Figure S6. A subject-wise permutation test with 10,000 permutations was used to estimate significance (see above for detailed description). For each predictor, the p-values were FDR-corrected across all ROIs.

For whole-brain analysis, we performed a similar analysis as for the functionally defined ROIs for volumetric cortex parcels defined by Glasser and colleagues[35]. We used 180 bilateral parcels combining the symmetrical parcels of the two hemispheres. First, we identified visually responsive voxels in each parcel using the intact > scrambled images contrast from the localizer, excluding voxels with t-values equal to or smaller than zero. Next, we selected the top 200 voxels from this contrast, matching the voxel number of the ROI analysis. For parcels

with fewer than 200 visually responsive voxels, we used all voxels with t-values above zero. The extraction of mean time courses, the computation of the ISCs, and the partial correlation of neural IS-RDMs with drawing-based predictor IS-RDMs for each parcel were performed identically to the ROI analysis. A permutation test provided p-values that were FDR-corrected across all parcels. Figure S7B shows all parcels significant after selection of the top 200 voxels when selecting different numbers of voxels to show that the results are robust across a range of voxel selection numbers. Lastly, for plotting the whole brain results, surface analogues of the significant parcels were highlighted on the right hemisphere of an inflated brain (see Figure 2C).

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG), grants KA4683/5-1 (project no. 518483074), KA4683/6-1 (project no. 536053998), and under Germany's Excellence Strategy (EXC 3066/1, "The Adaptive Mind", project no. 533717223). It was further supported by a European Research Council (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 funders. Neither the funders nor the granting authority can be held responsible for them.

MR imaging for this study was performed at the Bender Institute of Neuroimaging (BION) at Justus Liebig University Giessen, Germany.

We thank Malaika Grace Alphonsus and Kieran Pang for assistance with data collection and Lu-Chun Yeh and Susan Ajith for their feedback on the manuscript.

The authors declare that there are no competing interests.

Data and code availability

Data and code are available through the project's OSF page (https://osf.io/zjtwx/). The photos of participants' private rooms are not shared to protect their privacy.

References

1. Hartley, C. A. How do natural environments shape adaptive cognition across the lifespan? *Trends Cogn. Sci.* **26**, 1029–1030 (2022).

2. Russell, R., Duchaine, B. & Nakayama, K. Super-recognizers: People with extraordinary face recognition ability. *Psychon. Bull. Rev.* **16**, 252–257 (2009).

3. White, D. & Burton, A. M. Individual differences and the multidimensional nature of face perception. *Nat. Rev. Psychol.* **1**, 287–300 (2022).

4. Richler, J. J. *et al.* Individual differences in object recognition. *Psychol. Rev.* **126**, 226–251 (2019).

5. de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S. & Gegenfurtner, K. R. Individual differences in visual salience vary along semantic dimensions. *Proc. Natl. Acad. Sci.* **116**, 11687–11692 (2019).

6. Broda, M. D. & de Haas, B. Individual differences in human gaze behavior generalize from faces to objects. *Proc. Natl. Acad. Sci.* **121**, e2322149121 (2024).

7. Henderson, J. M. & Luke, S. G. Stable individual differences in saccadic eye movements during reading, pseudoreading, scene viewing, and scene search. *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 1390–1400 (2014).

8. Borovska, P. & de Haas, B. Individual gaze shapes diverging neural representations. *Proc. Natl. Acad. Sci.* **121**, e2405602121 (2024).

9. Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D. & Kriegeskorte, N. Unique semantic space in the brain of each beholder predicts perceived similarity. *Proc. Natl. Acad. Sci.* **111**, 14565–14570 (2014).

10. Lee, J. & Geng, J. J. Idiosyncratic Patterns of Representational Similarity in Prefrontal Cortex Predict Attentional Performance. *J. Neurosci. Off. J. Soc. Neurosci.* **37**, 1257–1268 (2017).

11. Lin, Q. & Lau, H. Individual differences in prefrontal coding of visual features. 2024.05.09.588948 Preprint at https://doi.org/10.1101/2024.05.09.588948 (2025).

12. Friston, K. A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 815–836 (2005).

13. Hohwy, J. *The Predictive Mind*. (OUP Oxford, 2013).

14. Kersten, D., Mamassian, P. & Yuille, A. Object perception as Bayesian inference. *Annu. Rev. Psychol.* **55**, 271–304 (2004).

15. Bainbridge, W. A., Hall, E. H. & Baker, C. I. Drawings of real-world scenes during free recall reveal detailed object and spatial information in memory. *Nat. Commun.* **10**, 5 (2019).

16. Fan, J. E., Bainbridge, W. A., Chamberlain, R. & Wammes, J. D. Drawing as a versatile cognitive tool. *Nat. Rev. Psychol.* **2**, 556–568 (2023).

17. Engeser, M. *et al.* Characterizing internal models of the visual environment. *Proc. R. Soc. B Biol. Sci.* **292**, 20250602 (2025).

18. Wang, G. *et al.* How do visual and conceptual factors predict the composition of typical scene drawings? 2025.09.15.676247 Preprint at https://doi.org/10.1101/2025.09.15.676247 (2025).

19. Wang, G., Foxwell, M. J., Cichy, R. M., Pitcher, D. & Kaiser, D. Individual differences in internal models explain idiosyncrasies in scene perception. *Cognition* **245**, 105723 (2024).

20. Wang, G., Chen, L., Cichy, R. M. & Kaiser, D. Enhanced and idiosyncratic neural representations of personally typical scenes. *Proc. R. Soc. B Biol. Sci.* **292**, 20250272 (2025).

21. Finn, E. S. *et al.* Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage* **215**, 116828 (2020).

22. Hasson, U. *et al.* Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Res.* **2**, 220–231 (2009).

23. Shen, Y. L. *et al.* Neural similarity predicts whether strangers become friends. *Nat. Hum. Behav.* **9**, 2285–2298 (2025).

24. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Preprint at https://doi.org/10.48550/arXiv.1409.1556 (2015).

25. Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. *IEEE Conf. Comput. Vis. Pattern Recognit.* 248–255 (2009) doi:10.1109/CVPR.2009.5206848.

26. de Lange, F. P., Heilbron, M. & Kok, P. How Do Expectations Shape Perception? *Trends Cogn. Sci.* **22**, 764–779 (2018).

27. Peelen, M. V., Berlot, E. & de Lange, F. P. Predictive processing of scenes and objects. *Nat. Rev. Psychol.* 1–14 (2023) doi:10.1038/s44159-023-00254-0.

28. Bartnik, C. G. *et al.* Representation of locomotive action affordances in human behavior, brains, and deep neural networks. *Proc. Natl. Acad. Sci.* **122**, e2414005122 (2025).

29. Greene, M. R., Baldassano, C., Esteva, A., Beck, D. M. & Fei-Fei, L. Visual scenes are categorized by function. *J. Exp. Psychol. Gen.* **145**, 82–94 (2016).

30. Kyle-Davidson, C., Zhou, E. Y., Walther, D. B., Bors, A. G. & Evans, K. K. Characterising and dissecting human perception of scene complexity. *Cognition* **231**, 105319 (2023).

31. Nagle, F. & Lavie, N. Predicting human complexity perception of real-world scenes. *R. Soc. Open Sci.* **7**, 191487 (2020).

32. Palmer, S. E., Schloss, K. B. & Sammartino, J. Visual Aesthetics and Human Preference. *Annu. Rev. Psychol.* **64**, 77–107 (2013).

33. Vessel, E. A., Maurer, N., Denker, A. H. & Starr, G. G. Stronger shared taste for natural aesthetic domains than for artifacts of human culture. *Cognition* **179**, 121–131 (2018).

34. Nastase, S. A., Gazzola, V., Hasson, U. & Keysers, C. Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* **14**, 667–685 (2019).

35. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).

36. Biederman, I. Perceiving Real-World Scenes. *Science* **177**, 77–80 (1972).

37. Caddigan, E., Choo, H., Fei-Fei, L. & Beck, D. M. Categorization influences detection: A perceptual advantage for representative exemplars of natural scene categories. *J. Vis.* **17**, 21 (2017).

38. Kaiser, D., Stein, T. & Peelen, M. V. Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proc. Natl. Acad. Sci.* **111**, 11217–11222 (2014).

39. Kaiser, D. & Cichy, R. M. Parts and Wholes in Scene Processing. *J. Cogn. Neurosci.* **34**, 4–15 (2021).

40. Torralbo, A. *et al.* Good Exemplars of Natural Scene Categories Elicit Clearer Patterns than Bad Exemplars but Not Greater BOLD Activity. *PLOS ONE* **8**, e58594 (2013).

41. Bar, M. Visual objects in context. *Nat. Rev. Neurosci.* **5**, 617–629 (2004).

42. Coutrot, A. *et al.* Entropy of city street networks linked to future spatial navigation ability. *Nature* **604**, 104–110 (2022).

43. Barrett, H. C. Towards a Cognitive Science of the Human: Cross-Cultural Approaches and Their Urgency. *Trends Cogn. Sci.* **24**, 620–638 (2020).

44. Kanai, R. *et al.* Information generation as a functional basis of consciousness. *Neurosci. Conscious.* **2019**, niz016 (2019).

45. Moutsiana, C. *et al.* Cortical idiosyncrasies predict the perception of object size. *Nat. Commun.* **7**, 12110 (2016).

46. Brandman, T. & Peelen, M. V. Interaction between Scene and Object Processing Revealed by Human fMRI and MEG Decoding. *J. Neurosci.* **37**, 7700–7710 (2017).

47. Grill-Spector, K. The neural basis of object perception. *Curr. Opin. Neurobiol.* **13**, 159–166 (2003).

48. Kaiser, D., Quek, G. L., Cichy, R. M. & Peelen, M. V. Object Vision in a Structured World. *Trends Cogn. Sci.* **23**, 672–685 (2019).

49. Malach, R. *et al.* Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl. Acad. Sci.* **92**, 8135–8139 (1995).

50. Kaiser, D. & Peelen, M. V. Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *NeuroImage* **169**, 334–341 (2018).

51. Epstein, R. A. & Baker, C. I. Scene Perception in the Human Brain. *Annu. Rev. Vis. Sci.* **5**, 373–397 (2019).

52. Henriksson, L., Mur, M. & Kriegeskorte, N. Rapid Invariant Encoding of Scene Layout in Human OPA. *Neuron* **103**, 161-171.e3 (2019).

53. Mueller, S. *et al.* Individual Variability in Functional Connectivity Architecture of the Human Brain. *Neuron* **77**, 586–595 (2013).

54. Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664–1671 (2015).

55. Castello, M. V. di O., Tour, T. D. la & Gallant, J. L. Individual differences shape conceptual representation in the brain. 2025.08.22.671848 Preprint at https://doi.org/10.1101/2025.08.22.671848 (2025).

56. Barbey, A. K., Koenigs, M. & Grafman, J. Dorsolateral prefrontal contributions to human working memory. *Cortex* **49**, 1195–1205 (2013).

57. Bar, M. The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* **11**, 280–289 (2007).

58. Bar, M. *et al.* Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci.* **103**, 449–454 (2006).

59. Boly, M. *et al.* Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence. *J. Neurosci.* **37**, 9603–9613 (2017).

60. Lamme, V. A. F. Towards a true neural stance on consciousness. *Trends Cogn. Sci.* **10**, 494–501 (2006).

61. Lau, H. Hitting the Right Note. in *In Consciousness we Trust: The Cognitive Neuroscience of Subjective Experience* (ed. Lau, H.) 0 (Oxford University Press, 2022). doi:10.1093/oso/9780198856771.003.0004.

62. Odegaard, B., Knight, R. T. & Lau, H. Should a Few Null Findings Falsify Prefrontal Theories of Conscious Perception? *J. Neurosci.* **37**, 9593–9602 (2017).

63. Wokke, M. E., Vandenbroucke, A. R. E., Scholte, H. S. & Lamme, V. A. F. Confuse Your Illusion: Feedback to Early Visual Cortex Contributes to Perceptual Completion. *Psychol. Sci.* **24**, 63–71 (2013).

64. Brascamp, J., Sterzer, P., Blake, R. & Knapen, T. Multistable Perception and the Role of the Frontoparietal Cortex in Perceptual Inference. *Annu. Rev. Psychol.* **69**, 77–103 (2018).

65. Clarke, A., Taylor, K. I. & Tyler, L. K. The Evolution of Meaning: Spatio-temporal Dynamics of Visual Object Recognition. *J. Cogn. Neurosci.* **23**, 1887–1899 (2011).

66. Clarke, A., Taylor, K. I., Devereux, B., Randall, B. & Tyler, L. K. From Perception to Conception: How Meaningful Objects Are Processed over Time. *Cereb. Cortex* **23**, 187–197 (2013).

67. Fyall, A. M., El-Shamayleh, Y., Choi, H., Shea-Brown, E. & Pasupathy, A. Dynamic representation of partially occluded objects in primate prefrontal and visual cortex. *eLife* **6**, e25784 (2017).

68. Miller, B. T., Vytlacil, J., Fegen, D., Pradhan, S. & D'Esposito, M. The Prefrontal Cortex Modulates Category Selectivity in Human Extrastriate Cortex. *J. Cogn. Neurosci.* **23**, 1–10 (2011).

69. Summerfield, C. *et al.* Predictive Codes for Forthcoming Perception in the Frontal Cortex. *Science* **314**, 1311–1314 (2006).

70. Ferrante, O. *et al.* Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature* **642**, 133–142 (2025).

71. Mashour, G. A., Roelfsema, P., Changeux, J.-P. & Dehaene, S. Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron* **105**, 776–798 (2020).

72. Sergent, C. *et al.* Bifurcation in brain dynamics reveals a signature of conscious processing independent of report. *Nat. Commun.* **12**, 1149 (2021).

73. Kawakita, G., Zeleznikow-Johnston, A., Takeda, K., Tsuchiya, N. & Oizumi, M. Is my "red" your "red"?: Evaluating structural correspondences between color similarity judgments using unsupervised alignment. *iScience* **28**, (2025).

74. Kleiner, M., Brainard, D. & Pelli, D. What's new in Psychtoolbox-3? (2007).

75. Xiao, J., Ehinger, K. A., Hays, J., Torralba, A. & Oliva, A. SUN Database: Exploring a Large Collection of Scene Categories. *Int. J. Comput. Vis.* **119**, 3–22 (2016).

76. Willenbockel, V. *et al.* Controlling low-level image properties: The SHINE toolbox. *Behav. Res. Methods* **42**, 671–684 (2010).

77. Kaiser, D. & Nyga, K. Tracking cortical representations of facial attractiveness using time-resolved representational similarity analysis. *Sci. Rep.* **10**, 16852 (2020).

78. Chen, G., Taylor, P. A., Shin, Y.-W., Reynolds, R. C. & Cox, R. W. Untangling the relatedness among correlations, Part II: Inter-subject correlation group analysis through linear mixed-effects modeling. *NeuroImage* **147**, 825–840 (2017).

79. Chen, G. *et al.* Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. *NeuroImage* **142**, 248–259 (2016).

80. Pajula, J. & Tohka, J. How Many Is Enough? Effect of Sample Size in Inter-Subject Correlation Analysis of fMRI. *Comput. Intell. Neurosci.* **2016**, 2094601 (2016).

81. Kiniklioglu, M., Pitcher, D. & Kaiser, D. Distinct Mechanisms of Surround Modulation for Dynamic Natural Scenes in Motion- and Scene-Selective Cortex. 2025.11.06.686956 Preprint at https://doi.org/10.1101/2025.11.06.686956 (2025).

82. Rosenke, M., van Hoof, R., van den Hurk, J., Grill-Spector, K. & Goebel, R. A Probabilistic Functional Atlas of Human Occipito-Temporal Visual Cortex. *Cereb. Cortex N. Y. NY* **31**, 603–619 (2020).

83. Julian, J. B., Fedorenko, E., Webster, J. & Kanwisher, N. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage* **60**, 2357–2364 (2012).

Supplementary Material

Table of contents:

Supplementary Experiment

To find the best pipeline to objectively and reproducibly quantify perceptual similarity between the drawing-based descriptors of participants' internal models, we conducted an experiment where human raters judged the pairwise similarities of a subset of the typical drawings based on their content, as well as their drawing style and quality. We then used a set of different DNN architectures and training regimes, as well as image preprocessing approaches (original drawings vs. generated photorealistic images), to compute pairwise similarities of the typical and control drawings of the same participants. Similarities of the typical drawings were used to predict similarity ratings of drawing content, while control drawings were used to predict drawing style and quality ratings.

Methods

Participants

For the drawing similarity rating task, 21 participants (referred to as raters to avoid confusion with the main experiments) were invited who had never participated in any drawing study before. One rater only conducted the first session and was excluded, leaving 20 raters (10 female, mean age 27.6, ± 3.0).

Procedures

The similarity rating task was split into two sessions. In each session, raters judged the similarity between typical bathroom or kitchen drawings derived from the first experiment (one half of the raters started with the kitchen drawings and the other half with bathrooms). In each

trial, raters saw one pair of drawings (each 9.5 × 11.8 degrees of visual angle, and 10.7 and 5.4 degrees eccentricity in x and y direction, respectively) and two rating scales from one (for not similar at all) to seven (very similar) for similarities in content and drawing style/quality, respectively. Instructions for content rating were: “Judge how similar the depicted rooms are. Thereby, take into consideration which objects are drawn and how they are configured”, and for drawing style/quality: “Judge how similar the drawings are with respect to the style the drawer chose or the general drawing ability of the drawer”. Raters were further asked to rate the two criteria isolated from each other. Ratings were provided by a mouse click of the respective number on the scale with no time restriction. For each category, raters performed 625 trials, including ten practice trials with drawings excluded from the analysis of the first experiment, all possible pairs of included drawings (35 drawings = 595 pairs), and 25 catch trials of five pairs that were each repeated five times to judge within-rater consistency.

Analysis

Before comparing the human ratings to similarity estimates of the DNNs, we evaluated the reliability within and between raters. All reliability measures were calculated for both scene categories separately and then averaged. To judge between-rater consistency, the median inter-rater correlation was calculated (Spearman correlations between the ratings of each pair of raters) separately for content and style/quality ratings. The reliability of ratings was assessed by estimating the split-half reliability of content and style/quality ratings using the average Spearman correlation over 1.000 iterations of randomly splitting the raters into two halves and correlating the mean ratings between the halves. To assess within-rater consistency, the intraclass correlation coefficient (ICC) was calculated across repeated pairs for each subject. ICC measures the degree of absolute agreement among measurements, while a value of one would indicate perfect agreement between repetitions, and zero indicates no agreement beyond chance.

Next, IS-RDMs were constructed using the content and style/quality rating averaged over all raters separately for the bathrooms and kitchens. The IS-RDMs of human raters were then compared to the IS-RDMs derived from DNN-pipelines as described in the Method section of Experiment 1, with two varying factors: (1) the DNN model used to extract visual features and (2) whether the raw drawings and or the generated, photorealistic images were used. We included a set of different DNN architectures and training materials, specifically: VGG-16 (trained on classifying ImageNet images, layer “fc7”), VGG-16 (trained on classifying Places365 images, layer “fc7”), GoogleNet (trained on classifying Places365 images, layer “inception_5b-output”), DINOv2 ViT-Base (last hidden state), and CLIP ViT-B/32 (final layer's class token). To test which DNN-pipeline predicts human ratings the best, the IS-RDMs of human drawing content ratings were correlated with the IS-RDMs based on typical drawings for different DNN-pipelines, and style/quality ratings were predicted using the copy control IS-RDMs. Statistical significance was evaluated using a permutation test as described in the Methods section of Experiment 1.

Results

The median inter-rater correlation was relatively low for both ratings (content: $r = 0.31$, style/quality: $r = 0.28$). This may be because there are only subtle differences between the drawings of the same category, and raters may interpret similarity in slightly different ways.

Nevertheless, the experiment still produced reliable results as indicated by the split-half reliability of r = 0.81 and r = 0.77, for content and style/quality ratings, respectively. The within-rater consistency indicated a moderate level of agreement across repeated image pairs, with an ICC of 0.45 for content ratings and 0.53 for style/ability ratings.

When comparing the rating of similarity in drawing content to DNN-based estimates of similarities, we found that the VGG-16 trained on ImageNet, including the intermediate step of generating photorealistic images from the drawings, could predict human ratings the best (Figure S7A; r = 0.42, p < 0.001) among all tested alternatives. Consequently, this DNN pipeline was applied in Experiment 1 and pre-registered for Experiment 2.

For predicting the style rating using the control drawings, the DINO model, which utilized raw drawings, revealed the best match (Figure S7B, r = 0.26, p < 0.001). However, the differences between DNN-pipelines were minor, and VGG-16 with generated images was only marginally lower (r = 0.24, p < 0.001). This indicates that our copy control condition successfully captures individual differences in the drawing style and quality when using VGG16 trained on ImageNet.

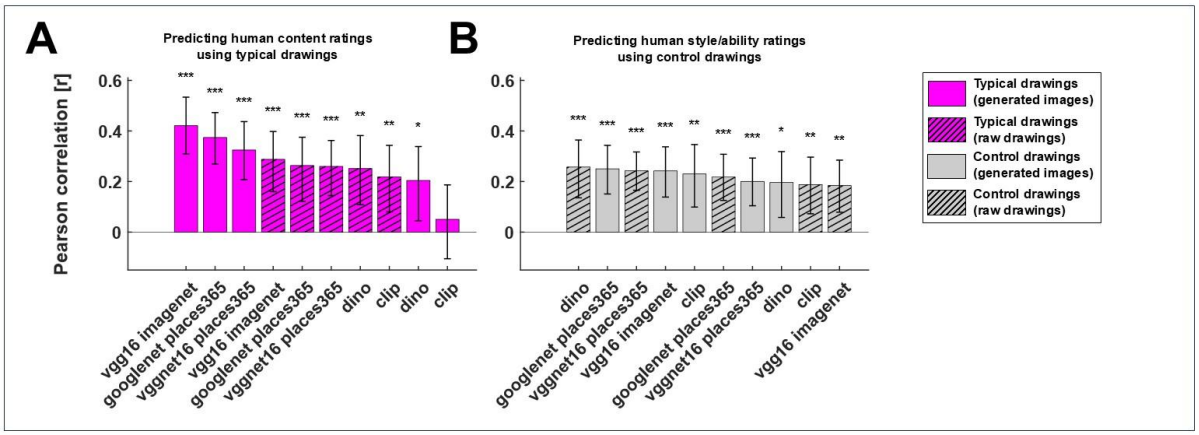

Figure S1 – Comparing different DNN pipelines for quantifying drawing similarity. *Similarities between a subset of the typical drawings produced in the main experiment were assessed by human raters separately for the content and the style/quality of the drawings. To objectively quantify the similarity between the same drawings, we used different DNN architectures and different training regimes, as well as original drawings (stripped bars) and photorealistic images generated using Draw3D (clean bars). Lastly, we tested which best predicts human ratings.* ***(A)*** *Sorted Pearson correlations of human similarity scores and DDN-derived similarities for the typical drawings (pink).* ***(B)*** *Sorted Pearson correlations of human style/quality similarity ratings of typical drawings and DDN-derived similarities for the copied drawings (grey) of the same participants. *** indicates p < 0.001, ** indicates p < 0.01, and * indicates p < 0.05, permutation tests, FDR-corrected for multiple comparison.*

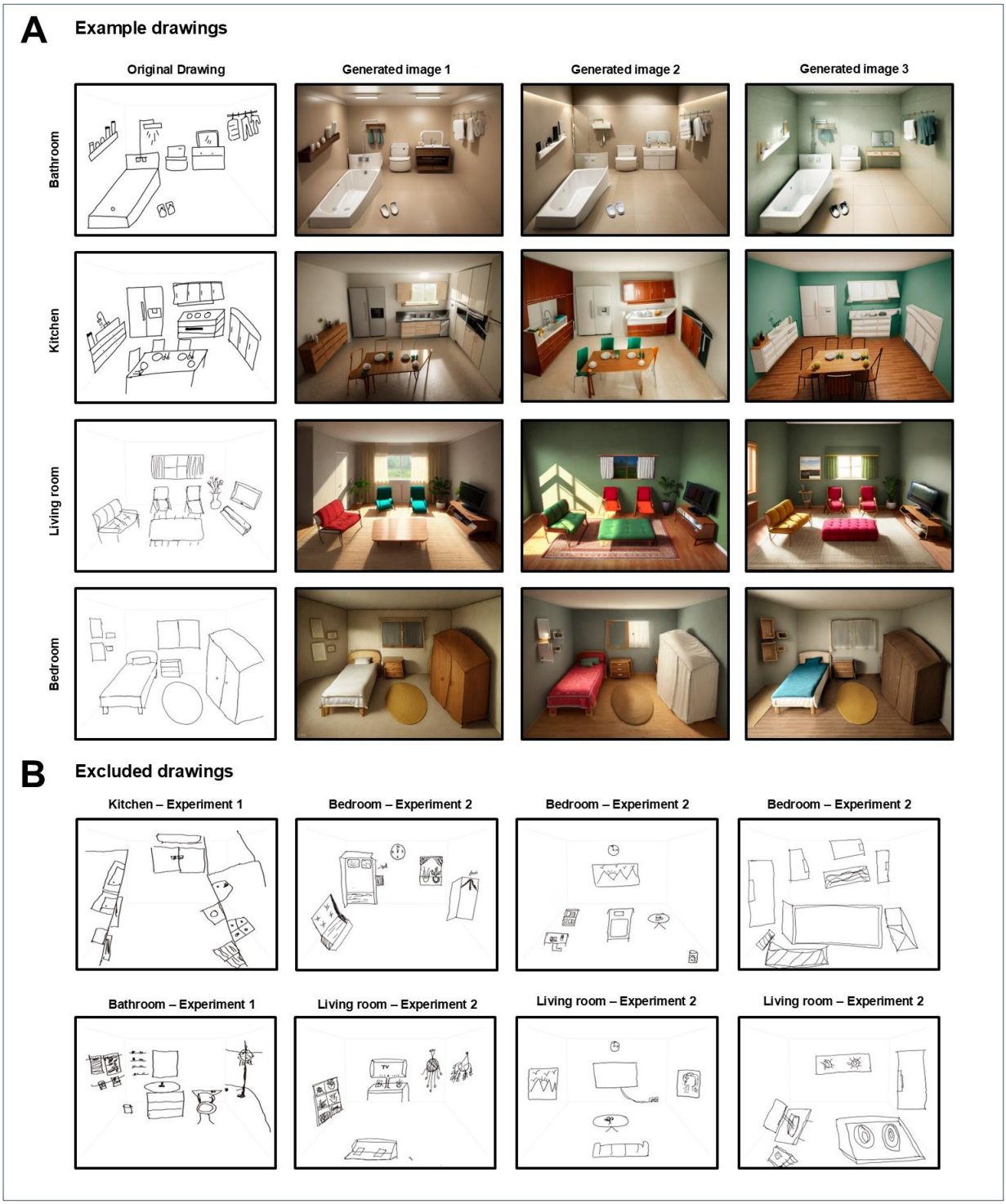

Figure S2 – Example drawings and excluded drawings. ***(A)*** *Representative drawings and photorealistic images generated from these drawings. Three images were generated for each drawing. The scene categories of the first behavioral and the fMRI experiment were kitchen and bathroom, and the second behavioral experiment used bedroom and living scenes. Note that all analyses were conducted on grey-scaled versions of the images.* ***(B)*** *Drawings from participants who were excluded because of insufficient drawing quality. One participant from Experiment 1 and three from Experiment 2 were excluded.*

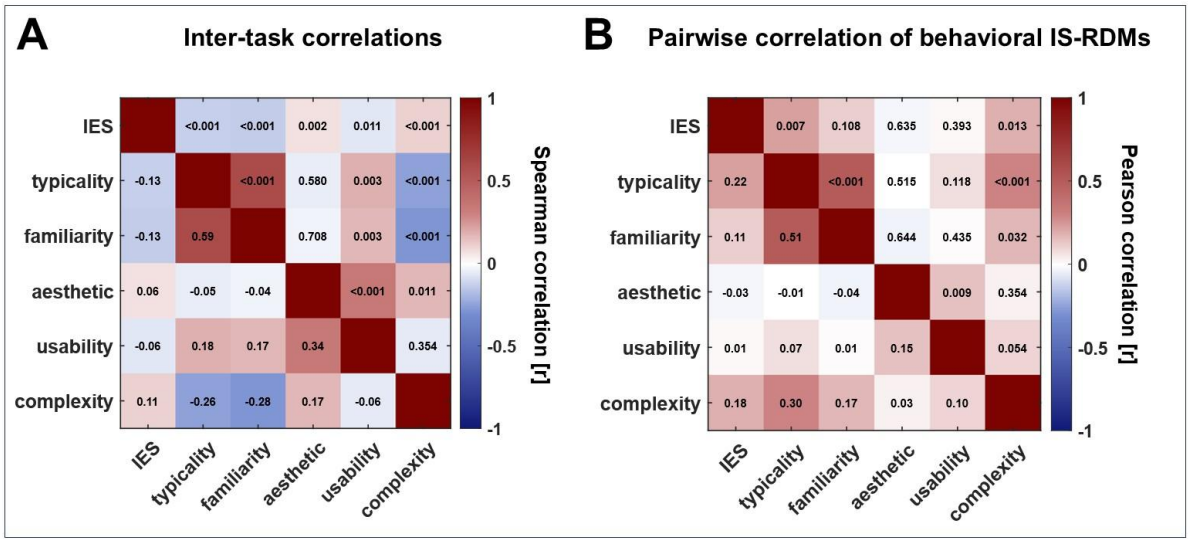

Figure S3 – Inter-task correlation of behavioral measures in Experiment 1. (A) *Inter-task correlation on the group level. For each participant, ratings or categorization performance were correlated across all images. The color and lower off-diagonal cell-labels display mean correlation coefficients across all participants. The upper off-diagonal cell-labels show the p-values resulting from a two-sided t-test against zero across all participants' correlation coefficients (FDR-corrected). Note that lower inverse efficiency scores (IES) indicate more efficient categorization performance; hence, negative correlations represent a positive relationship between categorization performance and the ratings (i.e., typical images are categorized more efficiently).* ***(B)*** *Correlation of inter-subject similarities across tasks. IS-RDMs for all behaviors were computed as described in the Method section. The color and lower off-diagonal cell-labels display correlation coefficients between each pair of measures. The upper off-diagonal cell-labels show the p-values resulting from permutation tests (FDR-corrected).*

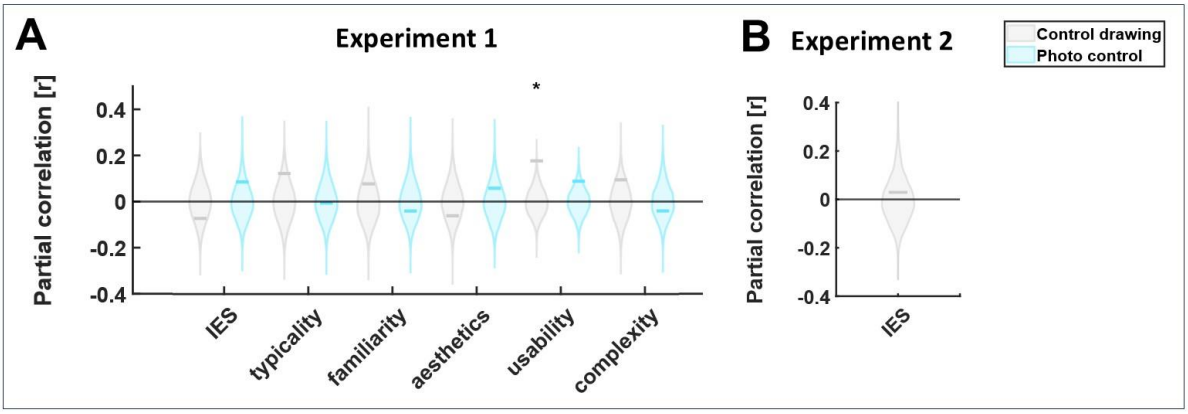

Figure S4 – Results for partial correlations of drawing and photo control predictors. *Behavioral IS-RDMs were compared with two predictor IS-RDMs based on copied drawings (grey) or provided photos (blue). In each analysis, partial correlations were used to control for typical drawings IS-RDM and the photo or copy control IS-RDMs, respectively.* ***(A)*** *Results of the partial correlation for the copy and photo controls in Experiment 1.* ***(B)*** *Results of control drawings correlation with categorization performance (IES) in Experiment 2. In the second experiment, no photos were provided. Violin plots show the empirical null distributions obtained from permutation tests, and the horizontal bars show the observed correlation coefficient. Asterisks indicate p<0.05 (permutation test, FDR-corrected for multiple comparisons where applicable).*

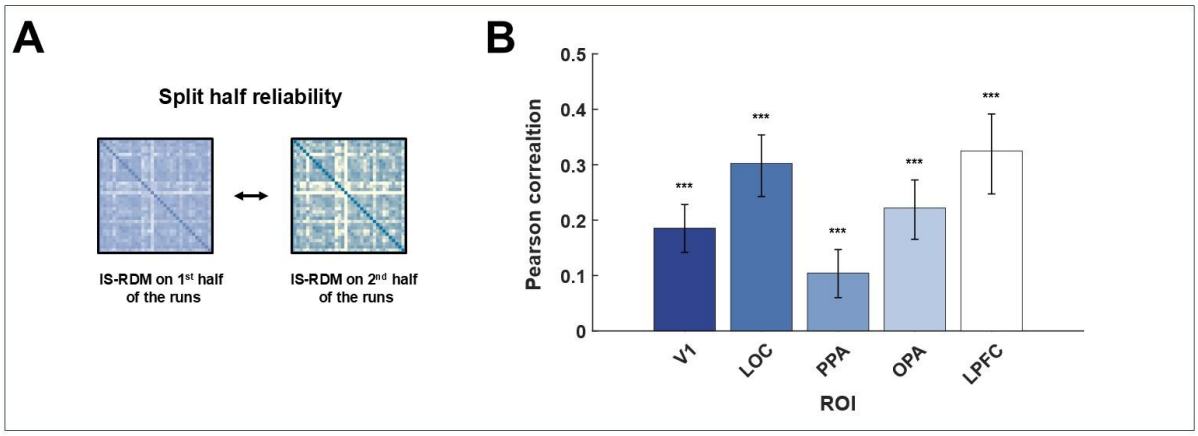

Figure S5 – Split-half reliability results for neural IS-RDMs. (A) *To assess the reliability of the inter-individual differences in the neural inter-subject correlation, we performed a split-half correlation analysis, correlating the IS-RDM built from two disjoint halves of runs. Positive correlations indicate reliable individual differences.* ***(B)*** *Results of split-half correlation for primary visual cortex (V1), lateral-occipital cortex (LOC), parahippocampal place area (PPA), occipital place area (OPA), and lateral prefrontal cortex (LPFC). Asterisks indicate p<0.001 (permutation test, FDR-corrected for multiple comparisons).*

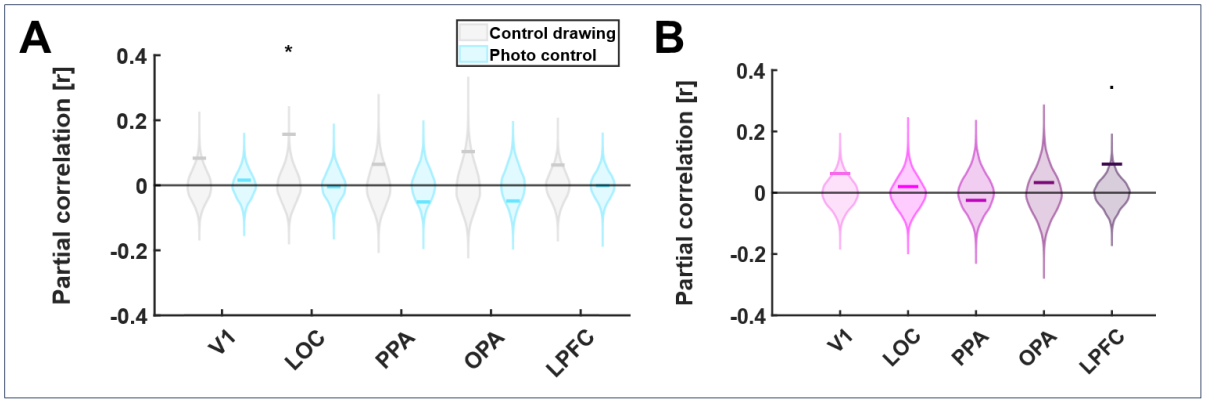

Figure S6 – Results for IS-RSA with control conditions and VGG16 trained on scene images. (A) *Partial correlation results of copied drawings (grey) and provided photos (light blue) with neural IS-RDMs (same as in Figure S3A).* ***(B)*** *IS-RSA results when using VGG16 trained on scene categorization (instead of object categorization) for primary visual cortex (V1), lateral-occipital cortex (LOC), parahippocampal place area (PPA), occipital place area (OPA), and lateral prefrontal cortex (LPFC). Violin plots show the empirical null distributions obtained from permutation tests, and the horizontal bars show the observed correlation coefficient. Asterisks indicate p<0.05 (permutation test, FDR-corrected for multiple comparisons).*

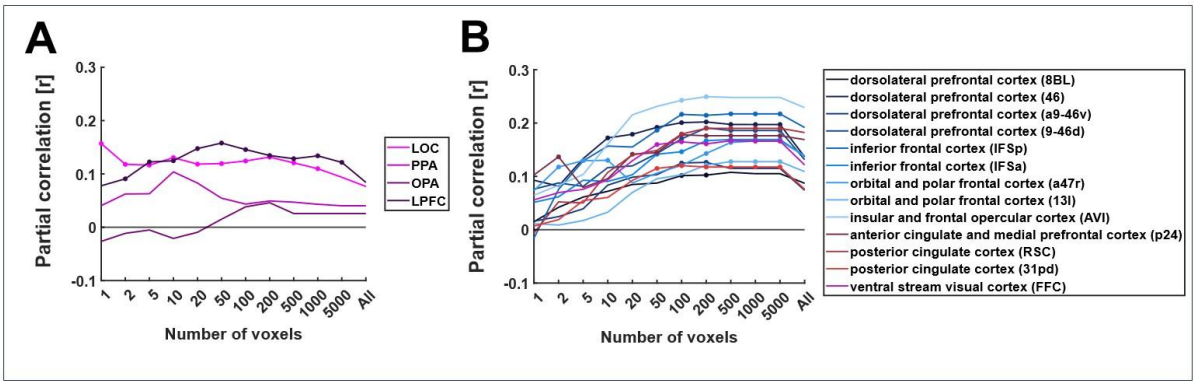

Figure S7 – fMRI results across different numbers of functionally selected voxels. *In the functionally defined ROIs and the parcels of the whole-brain analysis, the 200 voxels with the highest t-values in the first level GLM of the localizer were selected for computing ISC and partial correlation with drawing-based predictors. Here, the results are shown across different voxel numbers* ***(A)*** *for functionally defined ROIs and* ***(B)*** *for the brain parcels that were significant at the analysis using the top 200 voxels. For "All", no voxel selection was performed. Results were relatively stable across different voxel selection thresholds. Line plot displays the correlation coefficients and significance ($p < 0.05$, permutation test, FDR-corrected across ROIs/parcels with the same number of voxels) is indicated by dots.*

Fixed effect	**Experiment 1**			**Experiment 2**		
	χ^2	*df*	*p*	χ^2	*df*	*p*
Typical drawings	5.43	1	0.020	10.07	1	0.002
Control drawings	1.02	1	0.313	3.05	1	0.081
Photos	0.42	1	0.511	-	-	-
Typical drawings x control drawings	0.22	1	0.638	1.98	1	0.160
Typical drawings x photos	0.27	1	0.606	-	-	-
Control drawings x photos	1.42	1	0.233	-	-	-
Typical drawings x control drawings x photos	0.69	1	0.405	-	-	-

Table S1 – Full results from linear-mixed effect models. *An additional linear mixed-effect model with partially crossed random effects of subjects was employed to test how inter-subject similarities of categorization performance are predicted by inter-subject similarities in typical drawings, control drawings, and photos (photos were not collected in Experiment 2). The model included main effects and interactions of predictors (see Methods section of Experiment 2 for model formula). P values below 0.05 are considered significant.*

Parcel label	Parcel community	Community short	r	p-value (FDR)	p-value (uncorrected)
8BL	dorsolateral prefrontal cortex	dlPFC	0.10	0.044	0.003
46	dorsolateral prefrontal cortex	dlPFC	0.20	0.030	< 0.001
a9-46v	dorsolateral prefrontal cortex	dlPFC	0.13	0.044	0.003
9-46d	dorsolateral prefrontal cortex	dlPFC	0.19	0.030	< 0.001
IFSp	inferior frontal cortex	iFC	0.21	0.030	< 0.001
IFSa	inferior frontal cortex	iFC	0.17	0.032	0.002
a47r	orbital and polar frontal cortex	oFC and pFC	0.14	0.032	0.001
13l	orbital and polar frontal cortex	oFC and pFC	0.12	0.044	0.003
AVI	insular and frontal opercular cortex	insular and FO	0.25	0.044	0.003
p24	anterior cingulate and medial prefrontal cortex	aCC and mPFC	0.18	0.032	0.002
RSC	posterior cingulate cortex	pCC	0.19	0.032	0.001
31pd	posterior cingulate cortex	pCC	0.12	0.032	0.001
FFC	ventral stream visual cortex	v vis.	0.16	0.032	0.001

Table S2 – Full results of parcel-wise whole-brain analysis. *We performed a whole-brain analysis similarly to the analysis for the functionally defined ROIs, including 180 bilateral, volumetric cortex parcels. The table shows the parcel labels, parcel communities, an abbreviation of the community, the correlation coefficients (r), and the p-values before and after FDR-correction, for all 13 significant parcels after FDR-correction (p < 0.05). More information about the parcel can be found in Glasser et al.*[35]*.*