bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Activity in perirhinal and entorhinal cortex predicts perceived visual similarities among category exemplars with highest precision

Kayla M. Ferko [ab1], Anna Blumenthal [ad1], Chris B. Martin[e], Daria Proklova[a], Alexander Minos[a], Lisa M. Saksida[abc], Timothy J. Bussey[abc], Ali R. Khan[abfg], Stefan Köhler[ah]

[a] Brain and Mind Institute, University of Western Ontario, Canada

[b] Robarts Research Institute Schulich School of Medicine and Dentistry, University of Western Ontario, Canada

[c] Dept of Physiology and Pharmacology, University of Western Ontario, Canada

[d] Cervo Brain Research Center, University of Laval, Quebec, Canada

[e] Dept Psychology, Florida State University, USA

[f] School of Biomedical Engineering, University of Western Ontario, Canada

[g] Dept Medical Biophysics, University of Western Ontario, Canada

[h] Dept Psychology, University of Western Ontario, Canada

[1] Indicates shared first-authorship

[2] For correspondence: stefank@uwo.ca

Abstract

Vision neuroscience has made great strides in understanding the hierarchical organization of object representations along the ventral visual stream (VVS). How VVS representations capture fine-grained visual similarities between objects that observers subjectively perceive has received limited examination so far. In the current study, we addressed this question by focusing on perceived visual similarities among subordinate exemplars of real world-categories. We hypothesized that these perceived similarities are reflected with highest fidelity in neural activity patterns downstream from inferotemporal regions, namely in perirhinal and anterolateral entorhinal cortex in the medial temporal-lobe. To address this issue with fMRI, we administered a modified 1-Back task that required discrimination between category exemplars as well as categorization. Further, we obtained observer-specific ratings of perceived visual similarities, which predicted behavioural performance during scanning. As anticipated, we found that activity patterns

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

in perirhinal and anterolateral entorhinal cortex predicted the structure of perceived visual similarity relationships among category exemplars, including its observer-specific component, with higher precision than any other VVS region. Our findings provide new evidence that subjective aspects of object perception that rely on fine-grained visual differentiation are reflected with highest fidelity in the medial temporal lobe.

Introduction

The ability to perceive similarities and differences between objects plays an integral role in cognition and behaviour. Perceived similarities are important, for example, for categorizing a fruit at the grocery store as an apple rather than a pear. The appreciation of more fine-grained similarities between exemplars of a category also shapes behaviour, such as when comparing different apples in order to select one for purchase. Indeed, experimental work in psychology has confirmed that the similarity of objects influences performance in numerous behavioural contexts, including but not limited to categorization, object discrimination, recognition memory, and prediction (see Medin et al 1993; Goldstone & Son, 2012; Hebart et al 2020, for review). Yet, despite the well-established links to behaviour, how the brain represents these similarities between objects is only beginning to be understood. A central question that has received limited investigation so far is how fine-grained visual similarities that observers subjectively perceive among subordinate category exemplars map onto neural object representations. Answering this question can provide insight as to what brain regions provide the ‘read-out’ for such subjective reports. Moreover, this endeavor holds promise for understanding how differences in the way in which observers perceive their visual environment are reflected in variations in functional brain organization (Charest & Kriegeskorte, 2015).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Functional neuroimaging, combined with pattern analysis techniques, provides a powerful tool for examining the mapping between similarity relationships in visual perception of objects and similarities in corresponding neural representations (Kriegeskorte & Kievit, 2013). A significant body of research addressing this issue has focused on the characterization of category structure and other coarse object distinctions, such as animacy. Findings from this research indicate that activity patterns in a large expanse of the ventral visual stream (VVS), often referred to as inferotemporal cortex (IT) or ventral temporal cortex, capture much of this structure in the environment (Kriegeskorte et al., 2008; Connolly et al., 2012; Mur et al., 2013; Proklova et al., 2016; Cichy et al., 2019). For example, numerous studies have revealed a similarity-based clustering of response patterns for objects in IT that is tied to category membership (e.g., Kriegeskorte et al., 2008; Proklova et al., 2016; see Grill-Spector & Weiner., 2014 for review). In these studies, and in most related work, the primary focus has been on similarity in relation to object distinctions that are defined in objective terms, and on the characterization of neural representations that is shared by observers. As such, they do not address whether activity patterns in IT also capture similarity relationships among objects that characterize subjective aspects of visual perception that may vary across individuals' reports. When neural activity that corresponds to subjectively perceived visual similarities has been examined, extant research has mostly focused on specific object features, such as shape or size (Op de Beeck et al., 2008; Haushofer et al., 2008; Schwarzkopf et al., 2011; Moutsiana et al., 2016) rather than on similarities between complex real-world objects that differ from each other on multiple dimensions. A notable exception to this research trend is an fMRI study that focused on perceived similarities among select real-world objects that are personally meaningful (e.g., images of observers' own car, their own

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

bicycle; Charest et al., 2014), which demonstrated links between observer-specific perceived similarity and the similarity structure embedded in activity patterns in IT.

In order to reveal the mapping between fine-grained perceived visual similarities among category exemplars and similarities in neural activity, it may not be sufficient to focus on activity in IT but critical to consider VVS regions situated downstream on the medial surface of the temporal lobe. Perirhinal cortex (PrC), and the primary region to which it projects, i.e., lateral entorhinal cortex, are of particular interest when such perceived similarity relationships concern entire objects. The representational-hierarchical model of object processing in the VVS asserts that structures in the medial temporal lobe constitute the apex of its processing hierarchy (Murray & Bussey, 1999; Bussey & Saksida, 2007; Cowell, 2012; Kent et al., 2016). It proposes that there is a progressive integration of object features along the VVS that allows for increasing differentiation in the representation of objects from posterior to more anterior regions in the service of perception as well as other cognitive domains (e.g. recognition memory). Visual objects are thought to be represented in PrC in their most highly integrated form based on complex feature conjunctions (Murray & Bussey, 1999; Buckley & Gaffan, 2006; Bussey & Saksida, 2007; Graham, Barense, & Lee, 2010; Kent et al., 2016), which support perceptual discrimination of even highly similar exemplars from the same object category (e.g., O'Neil et al., 2009, 2013). Lateral entorhinal cortex (or its human homologue anterolateral entorhinal cortex; alErC; Maass et al. 2015; Olsen et al., 2017) has been suggested to extend this role in visual object discrimination through integration with additional spatial features (Connor & Knierim, 2017; Yeung et al., 2017). Taken together, these properties make PrC and alErC ideally suited for providing the read-out for

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

subjective reports of perceived visual similarity among the subordinate exemplars of object categories. Although more posterior VVS regions may also predict perceived similarity, the representational-hierarchical model asserts that they do not provide the same level of differentiation as regions at the apex in the medial temporal lobe. Moreover, they may also not capture those aspects of perceived similarity structure that are observer specific.

There is some evidence from human lesion studies suggesting that the integrity of regions in the medial temporal lobe is critical for perceiving the similarity among complex objects when it is high and objects cannot be easily discriminated from each other. This evidence comes from experiments that used variants of the oddity-discrimination task. In this task, participants view multiple objects (a minimum of three) on a computer screen and are asked to report the item that is least similar to the others. A finding documented in multiple reports (Buckley et al., 2001; Barense et al., 2007; Bartko et al., 2007; Inhoff et al., 2019; c.f., Stark and Squire, 2000; Levy et al., 2005; Hales et al., 2015) is that patients with medial temporal-lobe lesions that include PrC and alErC show deficits in oddity-discrimination tasks when complex objects with substantial feature overlap are compared, but not when oddity judgments require comparison of simple visual stimuli that can be distinguished based on a single feature such as colour or luminance. While the results of this lesion research have critically informed theoretical arguments that PrC plays a role in perceptual discrimination of objects (see Bonnen et al., 2021, for a recent computationally focused review), it is important to note that they do not provide a characterization of similarity structure of neural representations in PrC and alErC, nor a characterization of the transformation of representations from more posterior VVS regions

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

to these regions in the medial temporal lobe. Moreover, lesion findings do not speak to whether neural representations in the medial temporal lobe even capture the perceived similarity structure that is unique to individual observers.

In the current fMRI study, we tested the idea that the visual similarity structure among exemplars of real-world categories that observers subjectively perceive and report is predicted with higher precision by the similarity structure of neural activity in PrC and alErC than in the more posterior VVS regions traditionally considered in neuroscience investigations of visual object perception, including IT. During scanning, we administered a novel experimental task that required visual discrimination between consecutively presented exemplars from real-world categories as well as categorization (see Fig. 1). In addition, we obtained ratings of perceived visual similarities among these exemplars from each observer offline (see Fig. 2), as well as estimates of similarity derived from an influential computational model that describes objects at their intermediate visual feature level (HMAX, (Riesenhuber, Poggio, 1999; Cadieu et al., 2007; Serre, Olivia, Poggio, 2007; Khaligh-Razavi, Kriegeskorte, 2014). At the behavioural level, we found that discrimination performance was highly sensitive to fine-grained perceived similarity among exemplars, including those aspects that were observer-specific. Representational similarity analyses (RSA) of ultra-high resolution fMRI data revealed, in line with our hypotheses, that activation patterns in PrC and alErC predict this fine-grained structure within categories with higher precision than any other VVS region, and that they do so in a way that even captures those aspects of similarity structure that are unique to individual observers.

Results

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Perceived Visual Similarity Structure among Exemplars Varies across Observers

We used inverse multidimensional scaling (iMDS, Kriegeskorte & Mur, 2012) to create participant-specific models of perceived visual similarity for 4 exemplars from 10 different categories (see Fig. 1). Specifically, participants were instructed to arrange images of objects in a circular arena by placing those they perceived to be more visually similar closer together, and those they perceived to be less visually similar farther apart. Participants completed these arrangements offline, i.e., outside of the scanner, in two phases, with the first phase involving sorting of the full set of 40 objects in a single arrangement (Supplementary Fig. 1A). The second phase required sorting of exemplars within categories in 10 separate arrangements (one per category; Fig. 1A). Given our interest in representations that capture fine-grained object similarities within categories, our primary fMRI analyses relied on the similarity structures computed based on sorting in this second phase. The distances between all pairwise exemplars within each category were used to create a behaviour-based (i.e., subjective-report) representational dissimilarity matrix (RDM; Fig. 1B), which included a split of the range of similarities into three levels for sensitivity analyses in behaviour and neural activation patterns (see Methods for further detail). Examination of intersubject correlations of each participant's RDM and the mean of all other participants' RDMs (excluding their own) revealed a mean value of $r = 0.69$ (t-test $r>0$: $p<.001$). When we calculated an RDM for ratings averaged across participants, and compared it with an RDM derived from a computational model developed to capture objects at their intermediate visual feature level (HMAX, (Riesenhuber, Poggio, 1999; Cadieu et al., 2007; Serre, Olivia, Poggio, 2007; Khaligh-Razavi, Kriegeskorte, 2014), we also found a significant correlation ($r = 0.25$, $p< 0.001$),

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

suggesting that average ratings capture a shared component in the subjective ratings that relate to some objective image characteristics. Follow-up analyses that focused on this relationship in different subranges of similarity among exemplars (low, medium, high; see Figure 1B) revealed, however, that the estimates derived from the HMAX model were only significantly correlated with average ratings at low and medium levels (low $r = .45$; $p < .001$; medium $r = .24$; $p < .001$; high $r = .05$; $p > .05$). This pattern suggests that at their finest grain, judgments of similarity within categories rely on integrated object representations that (i) are not captured by intermediate feature level descriptions, and (ii) are observer-specific. Indeed, the range of intersubject correlations in reported similarity (calculated across exemplars and categories) that was present in our sample of participants provides direct evidence for variability across observers ($r = 0.54$-0.81; see Fig 1C, and Supplementary Fig. 1 for data on individual categories). There are also hints that this variability is most pronounced at the finest grain of judgements, as reflected in reduced intersubject correlations for high similarity exemplars (high $r = .64$; medium $r = .65$, low $r = .72$; high < low; $t(22) = 3.41$; $p < .001$). Critically, in a separate behavioural experiment conducted in another sample of participants, we found that individual differences in perceived visual similarity structure among exemplars are temporally stable observer characteristics (see Supplementary Fig. 2). We leveraged this variability across observers in subsequent analyses for assessing the precision of our fMRI results, examining whether the structure of neural representations in PrC and alErC even predict the perceived similarity structure that is unique to individual observers at its finest grain.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

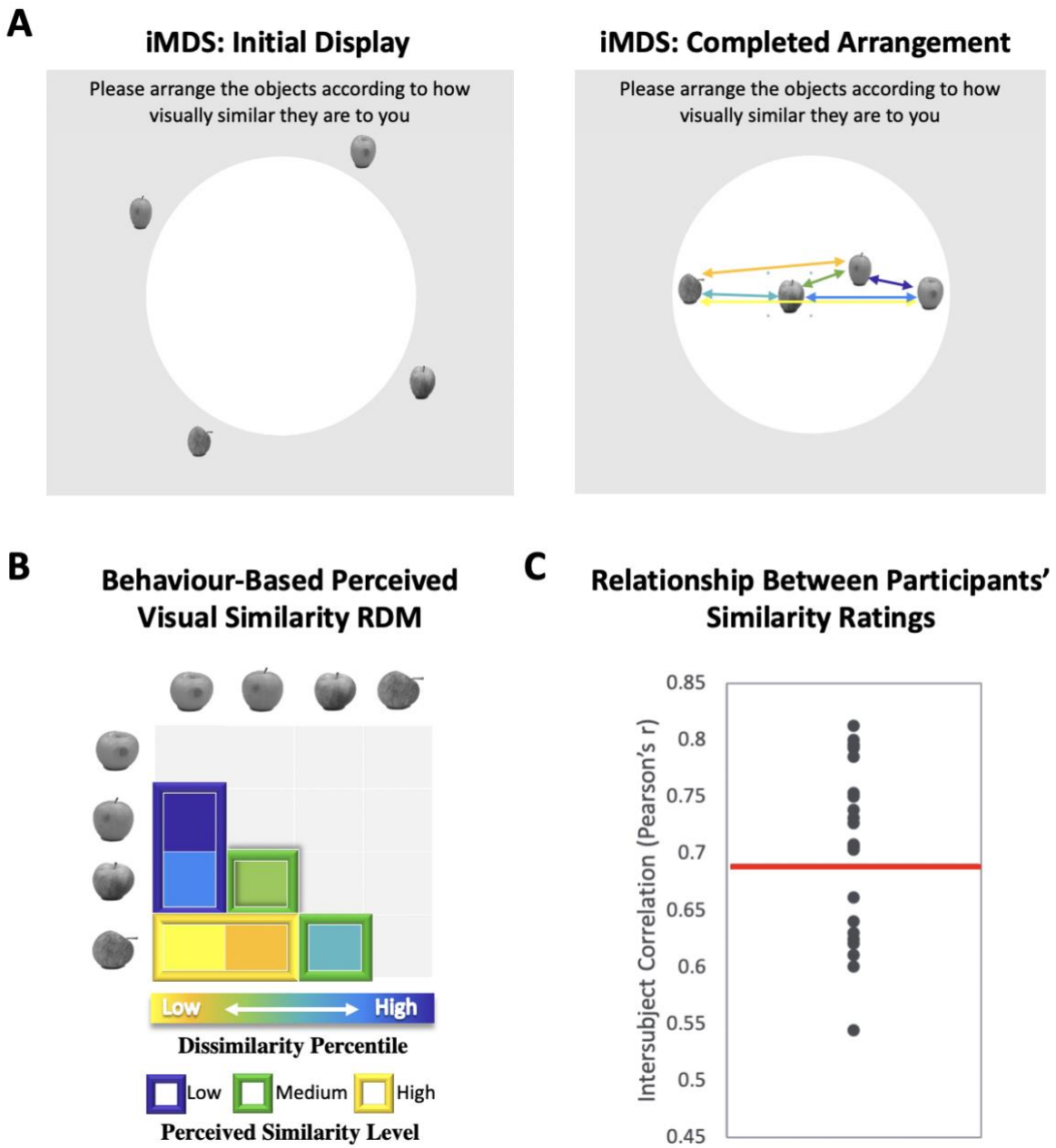


Figure 1. Perceived Visual Similarity ratings obtained offline with inverse multidimensional scaling (iMDS). **(A)** Task required placement of all exemplars from each category in circular arena, with distances reflecting perceived visual similarity. Arrows indicate the six pairwise distances used to compute Representational Dissimilarity Matrix (RDM). **(B)** Behaviour-based RDM computed using dissimilarity (1 – Pearson's r) and conversion to percentiles for individual observers. Only values below diagonal were included. Six distances (between 4 exemplars) per category were rank-ordered and grouped into three levels of similarity (low, middle, high; for more detail, Supp. Fig. 1). **(C)** Intersubject correlations for perceived similarity ratings across all exemplars and categories. Correlations were computed between each participant's RDM with the mean RDM (excluding the participant). Red horizontal line marks mean inter-subject correlation, with variability in perceived visual similarity structure across observers reflected in the range displayed.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Behavioural Discrimination Performance during Scanning is Sensitive to Observers' Perceived Visual Similarity Structure

Participants underwent ultra-high resolution fMRI scanning while completing a novel Category-Exemplar 1-Back Task designed to tax fine-grained visual object discrimination (see Fig. 2A). This task required responding to two different types of repetition, namely repetition of Same Exemplars or of different exemplars from the Same Category, across consecutive trials. Participants were asked to provide a button-press response when they noticed repetitions, with different buttons for each type of repetition. On all other trials, participants were not required to provide a response. Importantly, this task was designed to ensure that participants engaged in categorization, while also discriminating between exemplars within categories. Performance on the Category-Exemplar 1-Back Task was sensitive to perceived visual similarity between exemplars as reflected in observers' offline ratings and formalized in the behavioural RDMs with 3 different levels of similarity (Fig 2. B-C; see also Supplementary Table 1). Specifically, response errors on Same Category trials increased with increasing visual similarity (significant linear slope; $t(22) = 18.35$, $p < .0001$). Moreover, response times for correct responses on Same Category trials were positively correlated with perceived visual similarity level ($t(22) = 13.47$, $p < .0001$). Critically, task performance was also sensitive to the unique perceived similarity structure within categories expressed by observers. When we compared the influence of participant's own similarity ratings with that of others on behaviour (Fig. 2B-C), we found a significantly larger positive slope in error rate ($t(22) = 8.30$; $p<.0001$) and in response times ($t(22) = 9.68$; $p<.0001$) for participants' own ratings. This pattern of behaviour suggests that perceived visual similarity between exemplars influenced participants'

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

discrimination performance during scanning, and, that it did so in an observer-specific manner.

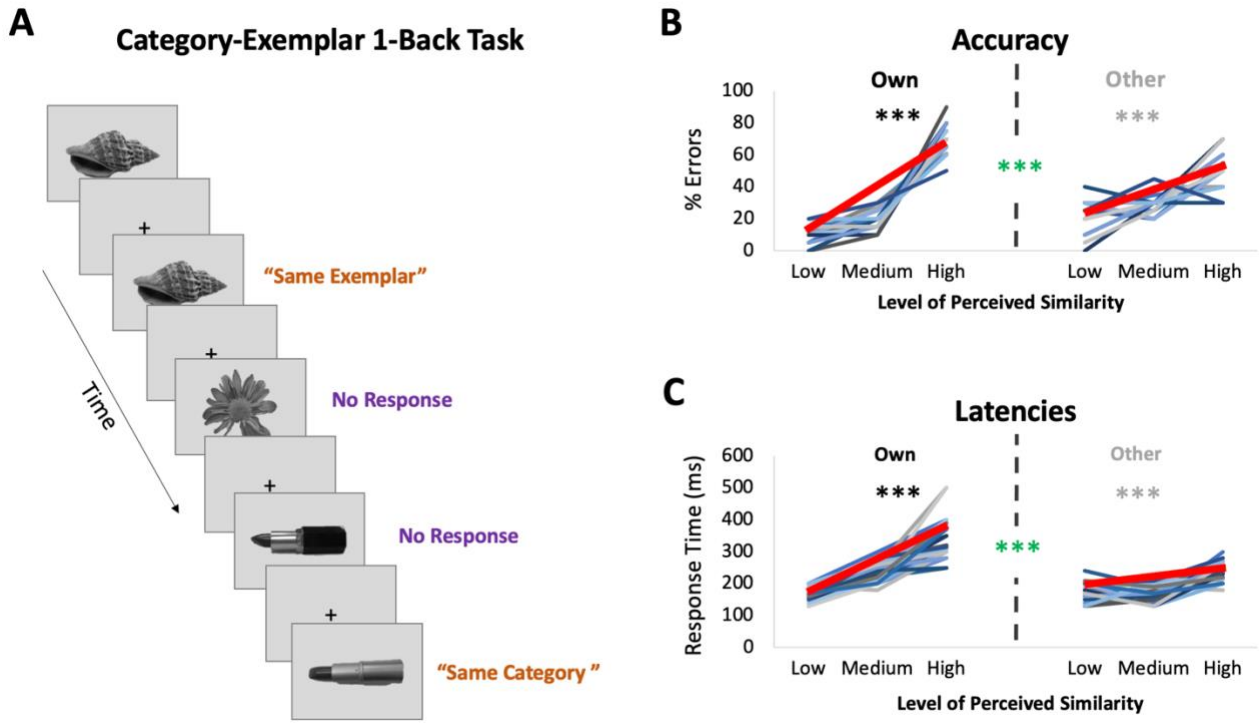


Figure 2. fMRI Task: Category-Exemplar 1-Back Task. **(A)** Images of objects depicting one of the 40 exemplars from 10 different categories were presented. Participants indicated repetitions on catch trials with two different button presses depending on whether the image was an exact repeat of the one previously presented (same exemplar, same category) or a repeat at the category level (different exemplar, same category). The majority of trials (75%) reflected no repetitions on either level and required no response. Only trials that required no response (non catch trials) were included in the fMRI analyses. **(B)** Percentage of errors that reflect Exact Repeat responses on Same Category trials as a function of perceived similarity (mean slope indicated with thick red line; for accuracy on all other trial types see Supplementary Materials Table 1). Own values reflect behavioral performance as a function of participants' own visual similarity ratings; Other values reflect performance based on other participants ratings (for a total of 22 iterations, which are then averaged). Error rate increased with increasing similarity as reflected in slopes (in black/grey ***P<.0001). Error rate was more sensitive to participants' own ratings as reflected in significantly higher slopes for the Own versus Other ratings (in green ***P<.0001). **(C)** Response times on correct Same Category trials as a function of perceived similarity (mean slope indicated with thick red line). Own and Other values calculated as in (B). Response times increased with increasing similarity and were more sensitive to participant's own than other participants' ratings. Results in (B,C) show that behavioural performance on 1-Back task is most sensitive to perceived similarity as reflected in participants' own ratings.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Patterns in Multiple VVS Regions Predict Perceived Visual Similarity Structure among Exemplars

To investigate whether the similarities between activation patterns in PrC and downstream alErC predict perceived similarities between exemplars in observers' reports, we employed anatomically defined ROIs and created participant-specific models of neural similarity between all 40 object exemplars on the no-response trials. In order to examine the anatomical specificity of our findings, we also created such models for ROIs in other VVS and medial temporal-lobe regions. Specifically, these ROIs included early visual cortex (EVC), lateral occipital cortex (LOC), posteromedial entorhinal cortex ErC (pmErC), parahippocampal cortex (PhC), and temporal pole (TP) for comparison (see Fig. 5 for visualization; note that LOC and PhC have typically been included in ROI definitions of IT in prior work; e.g., Charest et al., 2014; for direct comparison of results in IT and LOC see Supplementary Material Figure 5). Pairwise dissimilarities of neural patterns were employed to compute the brain-based RDMs (Fig. 3A-B); Pearson's correlations were calculated so as to examine whether these RDM's predicted participants' own behaviour-based RDMs that were derived from their offline reports of perceived similarity (Fig. 3C). Our analyses revealed that neural activation patterns in PrC and alErC did indeed correlate with participants' perceived visual similarity RDMs (Bonferroni-corrected $p<.01$). Patterns in other regions of the VVS (EVC; $p<.003$; LOC $p<.002$) were also significantly correlated with these behaviour-based RDMs. Critically, patterns in regions previously implicated in scene processing, specifically PhC and pmErC (Schultz et al., 2015; Maass et al., 2015; Schroder et al., 2015), did not predict the perceived similarity structure for objects (all $p > 0.5$). Having established that activity

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

patterns in multiple VVS regions predict perceived visual similarities between exemplars, we followed up on this finding by asking whether PrC and alErC capture these similarities with higher fidelity than earlier regions (EVC and LOC). Towards this end, we next examined whether these regions predict similarity structure even when exemplars only differ from each other in subtle ways.

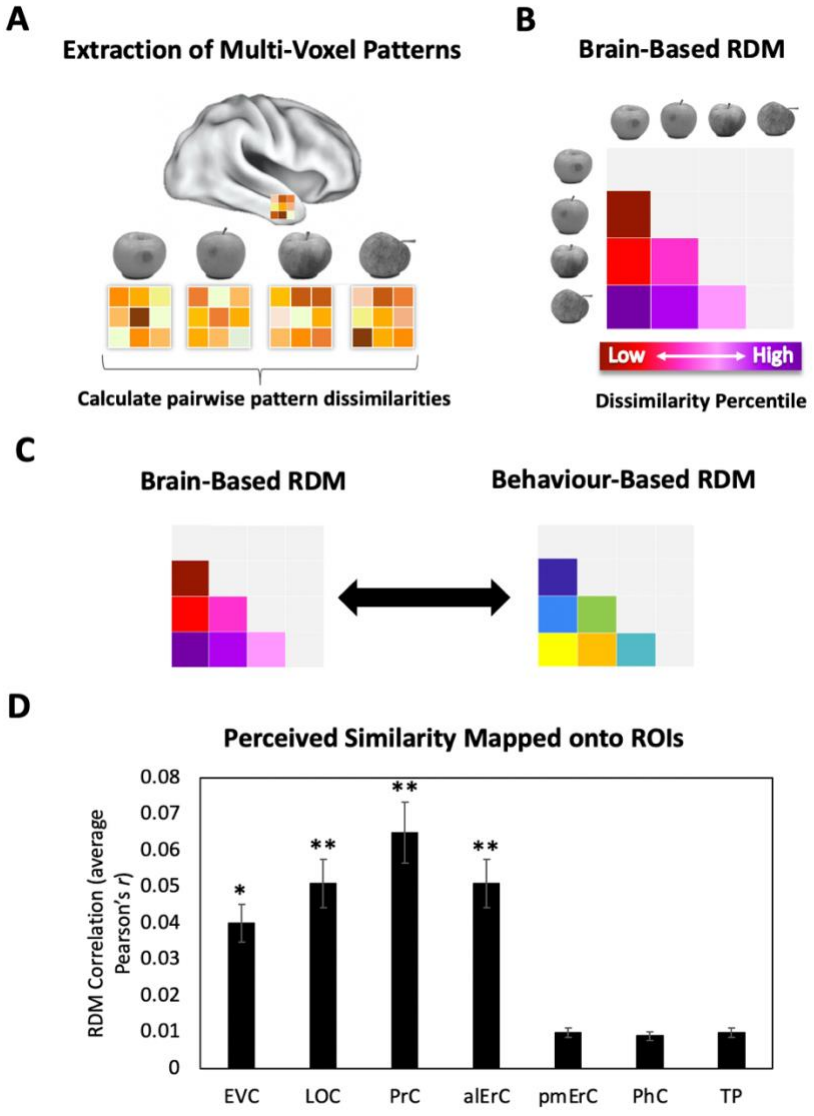


Figure 3. Brain-based representational dissimilarity matrices (RDMs) and their relationship to perceived visual similarity. **(A)** In each Region-of-Interest (ROI), mean multi-voxel activation patterns were calculated for every exemplar using the No Response trials in the Category-Exemplar 1-Back task. Pairwise pattern dissimilarities were computed as 1-Pearson's r. **(B)** Pairwise pattern dissimilarity percentiles were used to create participant-specific brain-based RDMs. **(C)** Brain-based RDMs were correlated with participants' own behaviour-based similarity

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

RDMs derived from their offline reports (black double arrows = within subject r). **(D)** Activation patterns in EVC, LOC, PrC, and alErC show significant correlations with participants' own perceived similarity ratings of objects (brain-based RDM X behaviour-based perceived similarity RDM within subjects; *P<.05, **P<.01 Bonferroni-corrected based on regions; Error bars represent SEM). EVC = Early Visual Cortex; LOC = Lateral Occipital Complex; PrC = Perirhinal Cortex; alErC = anterolateral entorhinal cortex; pmErC = posteromedial entorhinal cortex; PhC = Parahippocampal cortex; TP = Temporal Pole; see Fig. 5a for visualization).

PrC and alErC are the Only Regions Whose Patterns Predict Perceived Visual Similarity Structure among Exemplars when Similarity Is High

In this set of analyses, we examined the mapping between perceived similarity structure and neural activity patterns at a more fine-grained level within categories. Given prior evidence that PrC allows for the disambiguation of highly similar objects (Murray & Bussey, 1999; Buckley & Gaffan, 2006; Bussey & Saksida, 2007; Graham, Barense, & Lee, 2010; Kent, Hvoslef-Eide, Saksida, & Bussey, 2016), we anticipated that patterns in PrC, and possibly downstream alErC, would represent even the most subtle visual differences that observers perceive between exemplars, whereas earlier VVS regions would not. To address this issue with RSA in our stimulus set, participants' behaviour-RDMs, which were based on 6 pairwise distances between exemplars in each category, were divided into the three levels of similarity (low, medium, high; see Figure 1B and Supplementary Figure 1C for range comparison). Recall that our behavioural analyses revealed, as previously described, that discrimination performance on the Category-Exemplar 1-Back task was highly sensitive to these different levels of similarity. Figure 4 displays the results of our level-specific fMRI analyses, which were conducted for those regions whose activity patterns showed significant correlations with participants' full perceived visual similarity space between exemplars (as shown in Fig. 3D). Correlations

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

between perceived similarity structure and activity patterns in PrC and alErC were significant at all three levels of similarity (Bonferroni-corrected $p < .01$; Fig. 4A). In contrast, correlations for activity patterns in posterior VVS regions were significant only at the lowest level of perceived similarity (Bonferroni-corrected $p < .01$). This pattern of results cannot be attributed to differences in stability of activity patterns across levels of similarity for different regions; supplementary analyses revealed that stability was significant for all regions and did not interact with level of similarity (see Supplementary Figure 3). Direct comparison between regions also revealed that correlations in PrC and alErC were higher than in LOC and EVC at medium and high levels of perceived similarity (Bonferroni-corrected $p<.05$; Fig. 4A).

The described results suggest that object representations in posterior VVS regions may not have sufficient fidelity to allow for differentiation of exemplars that are perceived to be highly similar by observers, and that are most difficult to discriminate on our Category-Exemplar 1-Back task. We followed up on this idea with complementary classification analyses of our fMRI data using a linear support vector machine (Mur, Bandettini, & Kriegeskorte, 2009). These analyses were conducted so as to examine in which VVS regions activity patterns associated with exemplars of high perceived similarity would be sufficiently separable so as to allow for classification as distinct items. They confirmed that activity patterns associated with specific exemplars can indeed be successfully classified in PrC and alErC at higher levels of perceived similarity than in posterior VVS regions (Figure 4B for further detail).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

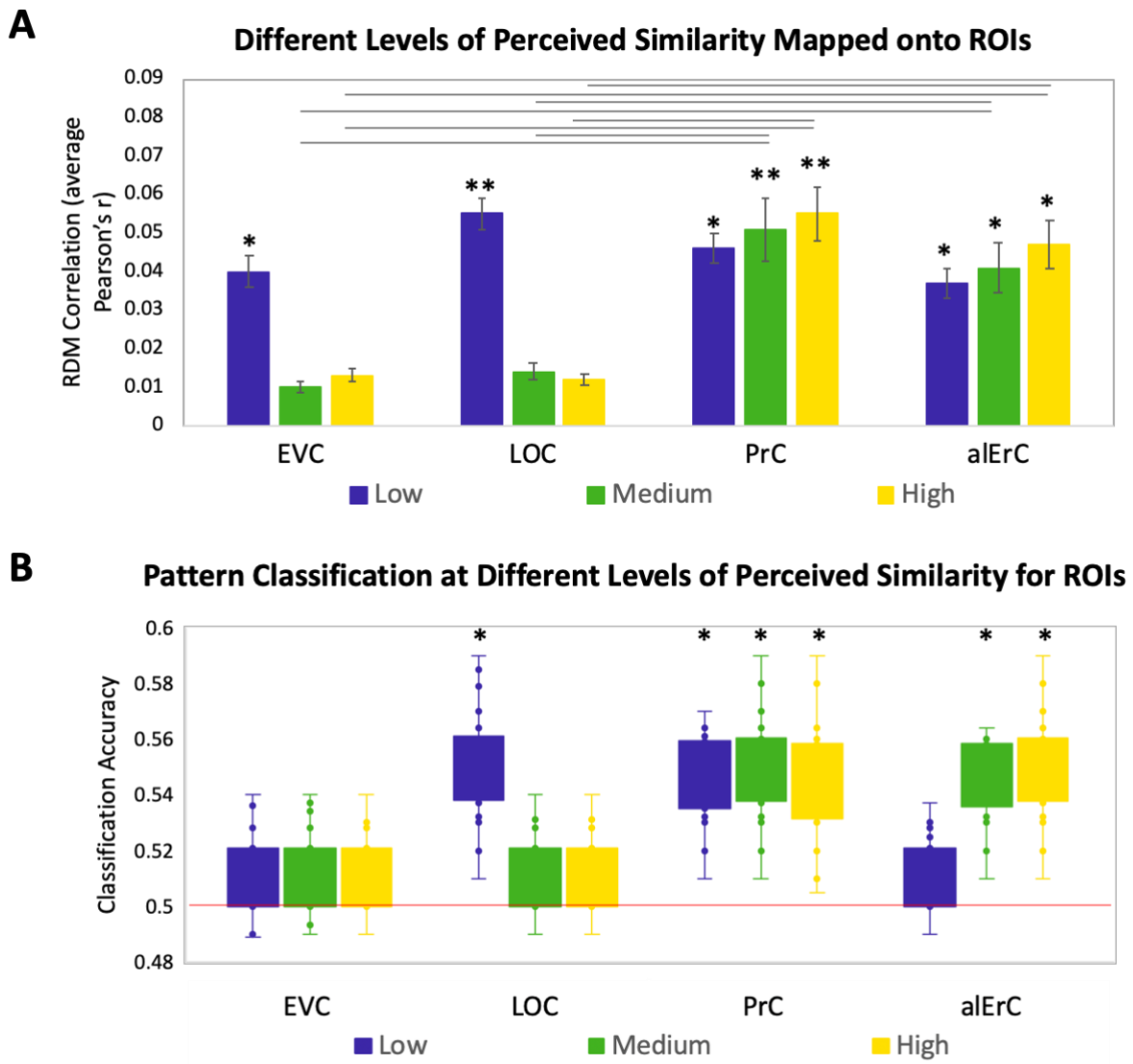


Figure 4. Relationship between brain RDMs and reports at different levels of perceived visual similarity for ROIs showing significant effects in Fig. 3d **(A)** Correlation of brain-based RDM and participants' own behaviour-based RDM at low, medium, and high levels of similarity. Only activation patterns PrC and alErC show significant correlation with ratings at middle and high levels of perceived similarity (*P<.01, Bonferroni-corrected for regions and levels). Correlations in PrC and alErC were significantly larger than those in EVC and LOC at the medium and high levels of perceived similarity (horizontal lines indicate P<.05). **(B)** Box and whisker plots for classification accuracy of neural activation patterns at each level of perceived similarity in different ROIs. We adopted a common classification approach using linear support vector classifier and leave-one-run-out cross-validation (Misaki, Kim, Bandettini, & Kriegeskorte, 2010). Results from one tailed t-tests to probe whether classifier performance was above chance indicate that patterns in LOC are distinguishable only at lowest level of perceived similarity within categories. PrC and alErC are only regions in which patterns are distinguishable at medium and high levels of perceived similarity. (*P<.05, Bonferroni-corrected).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The analyses presented so far only focused on specific regions of interest. To answer the question of whether PrC and alErC are the only regions whose activity patterns predict perceived visual similarity between exemplars within categories, we also conducted whole-volume searchlight-based RSA. As expected based on our ROI analyses, patterns in PrC and alErC, as well as in earlier VVS regions, showed a predictive relationship in these searchlight analyses when the full range of perceived visual similarity values between exemplars was considered; no regions in the scanned brain volume outside of the VVS exhibited this predictive relationship (Threshold- Free Cluster Enhancement corrected $p<.05$; Fig. 5C-D). Critically, our searchlight analyses revealed that PrC and alErC were indeed the only regions in the entire scanned brain volume whose patterns correlated with observer's reports when similarity between exemplars was perceived to be high, and objects were most difficult to discriminate on the Category-Exemplar 1-Back task (TFCE-corrected $p<.05$; Fig. 5E-F).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

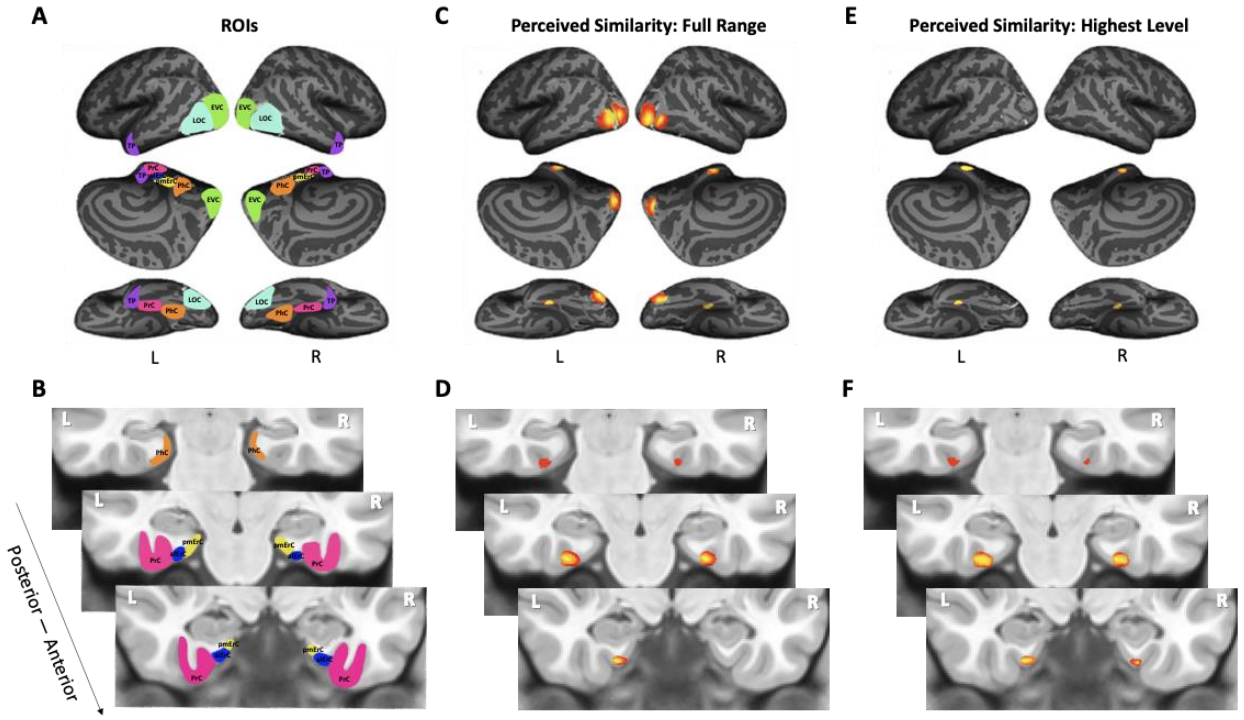


Figure 5. Visualization of ROIs and Results from Whole Volume Searchlight Analyses. **(A, B)** Visual depiction of ROIs. Early Visual Cortex (EVC = green), Lateral Occipital Complex (LOC = cyan), Perirhinal Cortex (PrC = pink), Parahippocampal Cortex (PhC = orange), Anterolateral Entorhinal Cortex (alErC = blue), Posteromedial Entorhinal Cortex (pmErC = yellow), Temporal Pole (TP = purple). **(C, D)** Cortical regions revealed with searchlight analysis in which brain-based RDMs were significantly correlated with behaviour-based perceived similarity RDMs across full range. Significant correlations were observed in PrC, alErC and more posterior VVS regions. **(E, F)** Cortical regions revealed with searchlight analysis in which brain-based RDMs were significantly correlated with behavioural RDMs at highest level of perceived similarity. Significant correlations were observed only in PrC and alErC. Maps are displayed with corrected statistical threshold of $p<.05$ at cluster level (using threshold-free cluster enhancement).

Patterns in Posterior VVS Regions Predict Similarity Structure that is Shared by Observers and Tied to Object Characteristics at the Intermediate Feature Level

Because our behavioural results point to shared as well as observer-specific components in the perceived similarity structure that participants reported, we conducted several additional sets of analyses, aiming to address which VVS regions would show activity patterns that might predict these different components. In this context, we considered that

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the perceived similarity structure among exemplars that is shared across observers for low and medium levels is tied to objective image characteristics at the intermediate feature level, as reflected in the significant correlation between averaged similarity ratings and estimates derived from the HMAX model in our data (see Behavioural Results above). Given the assertion of the representational hierarchical model that objects are represented in PrC and alErC based on complex feature conjunctions (rather than intermediate features), we predicted that the shared component of perceived similarity estimates of the HMAX model would predict activity patterns only in VVS regions posterior to PrC and alErC. Indeed, we found a significant predictive relationship of activity patterns in regions EVC and LOC (Bonferroni-corrected $p<.01$) but not in PrC and alErC (all Bonferroni-corrected $p>.05$; see Fig. 6A and B) for average perceived similarity. This significant relationship in posterior VVS regions was also only present at low (for EVC and LOC) and medium levels of similarity (for LOC; see Fig. 6B). Analyses that directly employed the estimates of similarity between object images obtained from the computational HMAX model (Riesenhuber, Poggio, 1999; Cadieu et al., 2007; Serre, Olivia, Poggio, 2007; Khaligh-Razavi, Kriegeskorte, 2014) revealed strikingly similar findings (see Fig 6B vs D). Activity patterns in EVC and LOC showed a significant correlation with the HMAX model at low and medium but not at the highest level of similarity; Bonferroni-corrected $p<.01$). Critically, activity patterns in PrC and alErC did not correlate with estimates from the HMAX model nor with average ratings at any level of object similarity (all Bonferroni-corrected $p>.05$; see Fig. 6C and D). Together, these results suggest that activity in VVS regions posterior to PrC and alErC capture the components of perceived visual similarity structure among exemplars that is shared by observers and that is closely related to object features at the intermediate feature level.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

At the same time, these neural representations in posterior VVS regions do not appear to allow for differentiation of exemplars at high levels of perceived similarity that tend to be observer-specific.

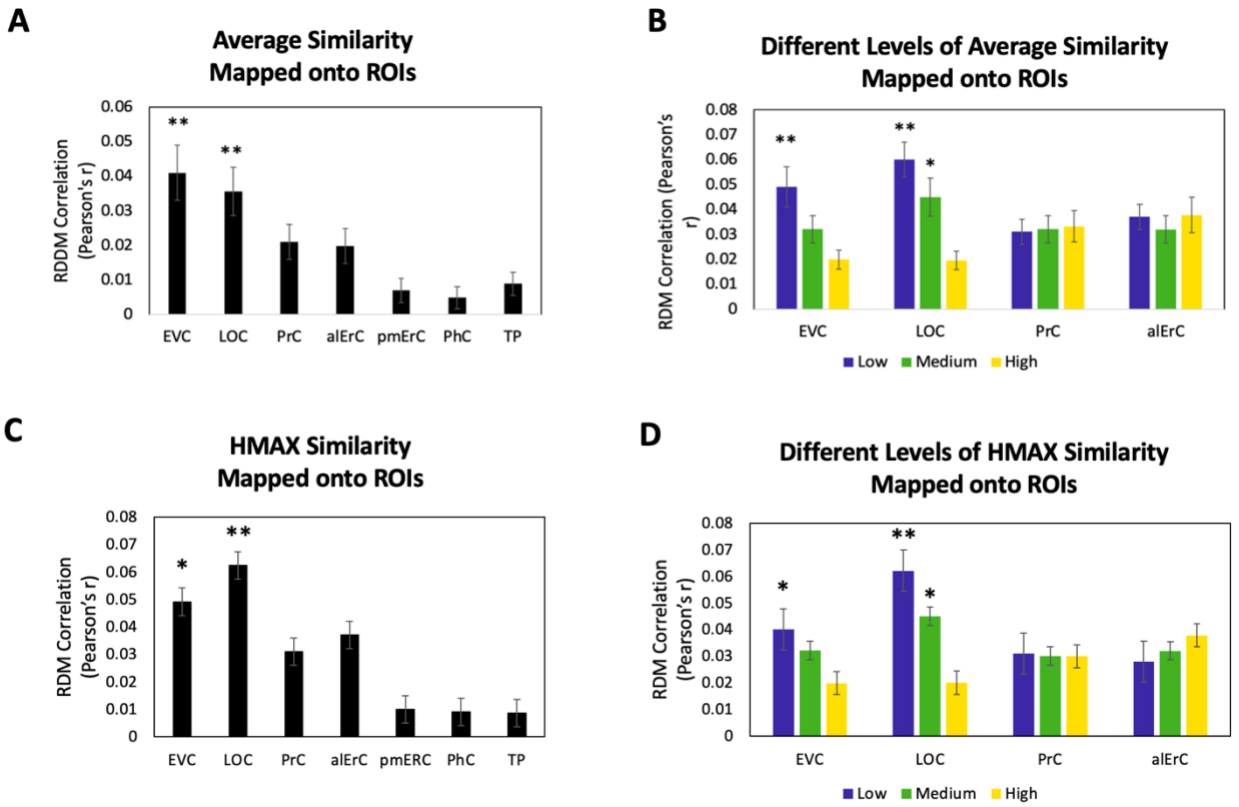


Figure 6. Brain-based representational dissimilarity matrices (RDMs) and their relationship to average perceived visual similarity. Brain-based RDMs were correlated with (A) the average behaviour-based similarity RDMs and (B) the different levels of average similarity RDMs; and with (C) the entire HMAX model (D) the different levels of similarity derived from the HMAX model Patterns in EVC and LOC show relationship to average whole perceived similarity ratings **P<.01, Bonferroni-corrected based on regions. EVC and LOC also show correlations to the average low, and LOC to the medium level of perceived similarity *P<.05, Bonferroni-corrected EVC = Early Visual Cortex; LOC = Lateral Occipital Complex; PrC = Perirhinal Cortex; alErC = anterolateral entorhinal cortex; pmErC = posteromedial entorhinal cortex; PhC = Parahippocampal cortex; TP = Temporal Pole)

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Patterns in PrC and alErC Predict Fine-Grained Perceived Visual Similarity Structure among Exemplars in an Observer-Specific Manner

In the final set of analyses we directly focused on the variability across participants' reports of similarity in order to address whether activity patterns in PrC and alErC even predict even those perceived similarities with high precision that are unique to individual observers. We reasoned that if neural patterns in a region represent the observer's *unique* perceived similarity structure, brain-behaviour correlations should be higher when calculated within rather than between participants (Fig. 7A black vs grey arrows; see Supplementary Table 2 for similar results revealed with a multiple regression approach). In other words, if there are observer-specific relationships, activity patterns should predict participants' own perceived similarity structure better than someone else's. Such analyses would reveal that interindividual differences are not only present in the perceptual reports of observers and in their discrimination performance, as demonstrated in our behavioural analyses, but also in corresponding neural representations in PrC and alErC. Indeed, initial analyses of our fMRI data demonstrated the presence of stable observer-specific activity patterns for the object exemplars probed in our study in all regions of interest (see Supplementary Materials Figure 4). The *i*-index introduced by Charest et al. (2014), which directly measures differences in correlations for the same (i.e., own) versus other observers, allowed us to examine which of these observer-specific activation patterns predict observer-specific structure in reports of perceived similarities. These analyses confirmed our expectation that the neural activation patterns in PrC and alErC predict observer-specific perceived visual similarity structure (Bonferroni-corrected P[within-participant r > between-participant r]<.01). In contrast,

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

activation patterns in posterior VVS regions, EVC and LOC (Bonferroni-corrected P[within-participant r> between-participant r]>.05) did not uniquely predict participants' own perceived similarity structure (Figure 7B). Not surprisingly, regions that did not predict participants' perceived similarity structure at all (pmErC, PhC, and TP; Figure 3D), also did not have significantly above zero *i*-indices (Bonferroni-corrected P[within-participant r > between-participant r]<.05). Critically, PrC and alErC showed significant Brain-Behaviour *i*-indices even when we restricted analyses to fine-grained differentiation, i.e., to the subrange of high levels of perceived similarity (Bonferroni-corrected P[within-participant r > between-participant r]<.05). Taken together, these results reveal that activity patterns in PrC and alErC even predict perceived similarities that are unique to individual observers, which are most prevalent in fine grained structure, with high precision .

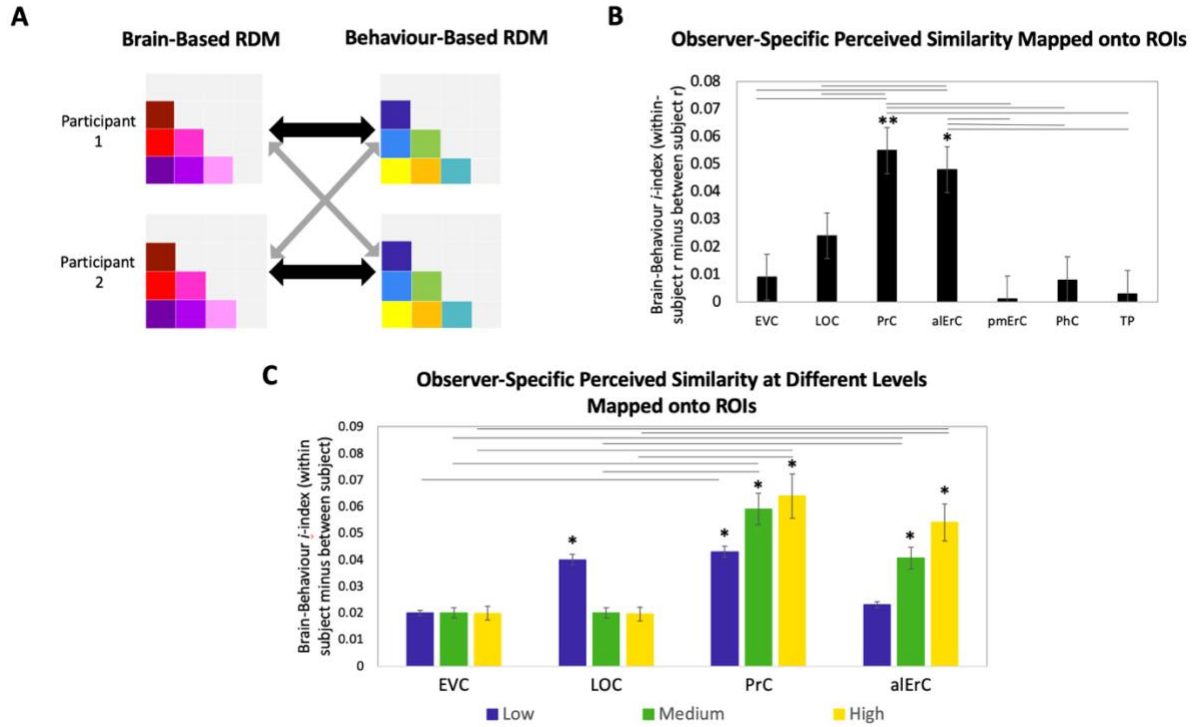


Figure 7. Brain-based representational dissimilarity matrices (RDMs) and their relationship to observer-specific perceived visual similarity. **(A)** Brain-based RDMs were correlated with (i)

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

participants' own behaviour-based similarity RDMs (black double arrows = within subject r) and (ii) other participants' behaviour-based similarity RDMs (grey arrows = between subject r) for comparison. **(B)** Patterns in PrC and alErC show relationship to perceived similarity ratings that are observer specific as reflected in Brain-Behaviour *i*-index (i.e., within minus between subject correlation; *P<.05, Bonferroni-corrected based on regions, with testing against a null-distribution created by randomizing subject-labels; Error bars represent SEM estimated based on randomization). PrC and alErC also show significant higher *i*-index than other regions as indicated with horizontal lines; *P<.05, Bonferroni-corrected. **(C)** Only patterns in PrC and alErC show relationship observer-specific perceived similarity ratings at the middle and high levels of perceived similarity. (*P<.01, Bonferroni-corrected for regions and levels). Correlations in PrC and alErC were significantly larger than those in EVC and LOC at the medium and high levels of perceived similarity (horizontal lines indicate P<.05). EVC = Early Visual Cortex; LOC = Lateral Occipital Complex; PrC = Perirhinal Cortex; alErC = anterolateral entorhinal cortex; pmErC = posteromedial entorhinal cortex; PhC = Parahippocampal cortex; TP = Temporal Pole; see Fig. 5a for visualization)

Discussion

Vision neuroscience has made great strides in understanding the hierarchical organization of object representations along the ventral visual stream (VVS). How VVS representations capture fine-grained differences between objects that observers subjectively perceive has received limited examination so far. In the current study, we addressed this question by focusing on perceived similarities among exemplars of real world-categories. Using a novel Category-Exemplar 1-Back Task, we found that visual discrimination performance is highly sensitive to the visual similarity structure that is reflected in observers' subjective reports. Combining this task with fMRI scanning at ultra-high resolution allowed us to show, in line with our general hypotheses, that activity patterns in PrC and alErC predict perceived visual similarities among exemplars with higher precision than any other VVS region, including prediction of those aspects of similarity structure that are unique to individual observers.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Research that has aimed to characterize the nature of object representations in human PrC with fMRI has shown that the degree of feature overlap between objects is captured by activation patterns in this region. Such a relationship has been revealed in multiple task contexts, with images of real-world objects and with words denoting such objects; moreover, it has been observed for feature overlap at the perceptual as well as the semantic level (Clarke and Tyler, 2014; Erez et al., 2016; Bruffaerts et al., 2013; Martin et al., 2018). These findings, in combination with work from neurophysiology in non-human animals and from computational modeling, have been interpreted to suggest that PrC integrates features of objects with complex conjunctive coding into representations of whole objects, and that the resulting conjunctive representations allow for differentiation of objects even when they are highly similar due to a high degree of feature overlap (Bussey & Saksida, 2002; Murray et al., 2007; Cowell et al., 2010). Indeed, it is this type of conjunctive coding that has motivated the central notion of the representational-hierarchical model of VVS organization that PrC, together with alErC (Connor & Knierim, 2017; Yeung et al., 2017), can be considered the pinnacle of the VVS object-processing hierarchy. With a metric that was rooted in participants' subjective reports of perceived visual similarity relationships and that we validated through examination of behavioural discrimination performance, our fMRI findings provide direct new evidence for increased differentiation of subordinate category exemplars in PrC and alErC, as compared to more posterior VVS regions.

Findings from lesion studies conducted with oddity-discrimination tasks support the idea that medial temporal-lobe structures downstream from IT play a critical role in processes required for the appreciation of fine-grained visual similarities between complex real-

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

world objects that are expressed in perceptual reports (see Bonnen et al., 2021, for review). Numerous studies conducted in humans and in other species have shown that performance on such tasks relies on the integrity of PrC when objects with high visual feature overlap must be judged (Buckley et al., 2001; Barense et al., 2007; Bartko et al., 2007; Inhoff et al., 2019; c.f., Stark and Squire, 2000; Levy et al., 2005; Hales et al., 2015). For example, Barense et al. (2007) compared performance on multiple visual oddity tasks between individuals with lesions in the medial temporal lobe that largely spared PrC and ErC, versus individuals with more widespread damage in the medial temporal lobes that included PrC and ErC. Most notably, individuals in the latter but not in the former group showed impairments in identifying the odd-one out item in sets of images of real-world objects that shared a high number of overlapping visual features. While the results of these prior lesion studies are compatible with the conclusions we draw in the current study, they do not allow for characterization of similarity structure of neural representations in PrC and alErC, and their direct comparison with representations in other medial temporal and posterior VVS regions, as provided here.

Indeed, the anatomical specificity of our fMRI findings in the medial temporal lobe is striking. While activity patterns that reflected the similarity structure among category exemplars were present in PrC and alErC, they were absent in medial-temporal regions that have previously been implicated in visual discrimination of scenes, specifically pmErC and PhC cortex (see Schultz et al., 2015, for review). This specificity is noteworthy in light of documented differences in functional connectivity between these regions that have been linked to object versus scene processing, with PrC being connected to alErC, and PhC being connected to pmErC, respectively (Maass et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2015; see Schröder et al., 2015 for more broadly distributed differences in functional connectivity between ErC subregions and other cortical structures).

The higher precision we observed for the representation of perceived similarity relationships in the medial temporal lobe, as compared to more posterior VVS regions, is of particular theoretical interest for the representational-hierarchical model of VVS organization (Murray & Bussey, 1999; Bussey & Saksida, 2007; Cowell, 2012; Kent et al., 2016). Most revealing, in this context, is the comparison between PrC and ErC versus LOC, a region that is part of the large swath of cortex that is often referred to as IT in neuroimaging research and that has been linked to processing of object shape in many prior fMRI studies (e.g., Grill-Spector et al., 2001; Kriegeskorte et al., 2008; Connolly et al., 2012; Mur et al., 2013; Proklova et al., 2016; Cichy et al., 2019). While activity patterns in LOC predicted some aspects of similarity structure among subordinate category exemplars in the current work, this relationship was observed at a coarser grain than in PrC and alErC; it only held when perceived visual similarity was low, and when performance in behavioural discrimination revealed that objects were easily distinguishable. Indeed, complementary pattern classification analyses revealed that activity patterns associated with exemplars of high perceived similarity were not sufficiently separable in LOC so as to allow for classification as distinct items. By contrast, this classification could be successfully performed based on activity patterns in PrC and alErC. Indeed, our searchlight analyses showed that these two regions in the medial temporal lobe were the only ones in which activity was related to perceived visual similarity at a fine-grained level.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Our analyses of the relationship between activity patterns and aspects of perceived similarity among exemplars that are tied to objective image characteristics, as estimated by the computational HMAX model, offer support for the central claim of the representational-hierarchical model that the transformation of object representations from IT (specifically, LOC) to medial temporal-lobe structures involves further integration. Notably, the component of perceived similarity structure that was shared by observers showed a statistical relationship to the estimates of the HMAX model, which describes objects at the intermediate feature level and which has been linked to LOC representations in prior work (Riesenhuber, Poggio, 1999; Cadieu et al., 2007; Serre, Olivia, Poggio, 2007; but see Khaligh-Razavi, Kriegeskorte, 2014, Kubilius et al., 2016 for limitations as compared to deep convolutional neural network models). In the present study, we also found that HMAX estimates of the similarity among the exemplars we employed were correlated with activity patterns in LOC, but only at low and medium levels of similarity. The fine-grained structure among exemplars that led to the largest number of confusion errors in behavioural discrimination on the Category-Exemplar 1-Back Task during scanning was predominantly observer-specific, and this fine-grained observer-specific structure was solely predicted by activity patterns in PrC and alErC. The coding of objects in PrC and alErC as fully integrated entities based on complex feature conjunctions, as proposed by the representational-hierarchical model, arguably affords the flexibility that is required to capture the fine-grained differentiation among exemplars that intermediate feature-level descriptions of visual objects cannot provide, and that were characterized by a high degree of observer specificity in the current data.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Our study was not designed to directly address what factors might drive the variability in perceived similarity structure among highly similar exemplars across observers that were present in subjective reports, their discrimination performance, and corresponding activity patterns in PrC and alErC. Prior evidence from fMRI research on neural representations in other VVS regions suggests that object familiarity may play an important role. Charest et al. (2014) revealed observer-specific effects in the similarity structure of activity patterns in IT that were tied to participants' reports for highly familiar real-world objects with unique personal meaning (e.g., images of observers' own car, their own bicycle). Notably, this observer-specific mapping between similarity in activity patterns and reports was not present for unfamiliar objects. There is also evidence from behavioural training studies indicating that prior experience with categories has an impact on how the similarity among its exemplars is perceived. In a recent study by Collins and Behrmann (2020), for example, it was shown that just a few days of repeated exposure can lead to increased differentiation among exemplars, and that these changes are most pronounced at the level of fine-grained similarity structure when observers have had some prior experience with the category in question. This change in similarity structure based on training occurred in the absence of any apparent opportunity to gain new sematic knowledge about the exemplars in question, suggesting it could reflect an increase in perceptual expertise. Indeed in recent behavioural research from our lab we have found that the degree of self-reported exposure to real-world object categories, but not corresponding semantic knowledge, predicts observers' perceived visual similarity structure among their exemplars, and that this relationship is most notable at the level of fine-grained structure (Minos et al, 2021). It is possible that the observer specificity in fine-grained differentiation among exemplars that was predicted by activity patterns in

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

PrC and alErC in the current study is tied to similar factors at work, and reflects interindividual differences in perceptual expertise across observers and categories. Although speculative at present, such an account would be in line with a large body of evidence revealing plasticity in object representations in these medial-temporal lobe structures (see Banks et al., 2014, for review). This account can be directly tested with training paradigms that target specific real-world object categories in future fMRI research. Regardless of the outcome of such future research, the current findings highlight the critical value of probing the subjective appreciation of visual object similarities, and their variability across observers, for a complete understanding of the transformation of neural representations from posterior regions to those at the apex of the VVS.

Material and Methods

Participants

A total of 29 participants completed the perceived similarity inverse multi-dimensional scaling arrangement task and fMRI experiment (12 females; age range = 18-35 years old; mean age = 24.2 years). All participants were right-handed, fluent in English, and had no known history of psychiatric or neurological disorders. Three participants were removed due to excessive head motion above the cut-off of 0.8 mm of framewise displacement, one participant were removed due to behavioural performance accuracy 2 SD below the average on the fMRI task, and two participants were removed due to temporal signal-to-noise ratio 2 SD below average. Therefore, 23 participants were included in the final analyses. All participants gave informed consent, were debriefed, and received monetary

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

compensation upon completion of the experiment. This study was conducted with Western's Human Research Ethics Board approval.

Stimuli

In order to investigate high-level object representations, we selected stimuli with varying levels of perceived visual similarity from the Migo Normative Database (Migo et al., 2013). We used 40 greyscale images of objects from 10 categories (Supplementary Figure 1A). Each category was made up of four exemplars that all shared the same name (e.g., apple, lipstick, stapler). Based on findings from a pilot study (n = 40), stimuli with similar perceived similarity ratings to the normative findings in the database were selected.

Modeling of Perceived Visual Similarity Structure

In order to obtain observer-specific models of perceived visual similarity structure for our stimuli, participants provided reports of perceived similarity between all stimuli on a computer outside of the scanner prior to scanning. Participants were seated in front of a monitor and completed a modified version of the inverse multi-dimensional scaling task (iMDS; Kriegeskorte & Mur, 2012). Specifically, participants were asked to drag-and-drop images into a white circle (i.e., arena), and arrange them according to perceived visual similarity (Kriegeskorte & Mur, 2012; see Figure 2A). Objects perceived to be more visually similar were placed closer together and objects perceived to be less visually similar were placed further apart. The iMDS task consisted of two phases. In the first phase, participants arranged all 40 stimuli according to perceived visual similarity (Supplementary Figure 3A). All unique pairwise distances were converted to dissimilarity

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

percentile and used to compute observer-specific models of the between-category similarity space. In the second phase, participants completed 10 category-specific trials in which they sorted four exemplars from the same category according to their perceived visual similarity. They were instructed to use the entire space within the circle, and make sure they compared each stimulus to every other stimulus. A MATLAB-based toolbox was then used to calculate distances between each pair of exemplars, and to convert these distances to dissimilarity percentiles (Kriegeskorte & Mur, 2012). These dissimilarity percentiles were then used to create observer-specific behaviour-based RDMs that represented each observers' perceived similarity space at the exemplar level. The behaviour-based RDMs for the entire range included 6 dissimilarity percentiles for each of the 10 categories (i.e., 6x10=60 dissimilarity percentiles)

These behaviour-based RDMs, which captured the full range of perceived similarity, were used to create 3 behaviour-based RDMs to reflect 3 levels of perceived similarity (low, medium, high). The 6 pairwise distances (expressed as dissimilarity percentile) per category were sorted into the two largest, two medium, and two smallest dissimilarities to create behaviour-based RDMs for low, medium, and high perceived visual similarity, respectively. These behaviour-based RDMs for each of the levels included 2 dissimilarity percentiles for each of the 10 categories (i.e., 2x10=20 dissimilarity percentiles). In order to ensure that the different levels of perceived similarity were non-overlapping, any values that did not allow for at least 0.1 dissimilarity percentile between each of the successive levels (i.e., high-middle, middle-low) was excluded. The range of dissimilarity percentiles did not differ significantly between the different levels of perceived similarity ($p>.05$; Supplementary Figure 1C).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Modelling of Objective Visual Similarity: HMAX Model

We obtained estimates of similarities that were derived from a computational model, HMAX (Riesenhuber & Poggio, 1999; Cadieu et al., 2007; Serre, Olivia, Poggio, 2007; Khaligh-Razavi, Kriegeskorte, 2014), developed to describe objects at their intermediate visual feature level, including shape. In this biologically inspired model, simple cell layers, akin to V1 cells, detect local orientation using Gabor filters. These orientation signals are pooled in complex cell layers to extract global features. In this way, HMAX is designed as a 4-layer hierarchical feed-forward structure similar to that described in the Ventral Visual Stream. The output layer of HMAX captures an object's shape over activation patterns and has been shown to correspond to activation patterns in IT cortex and LOC (Khaligh-Razavi, Kriegeskorte, 2014). We used Matlab implementation of HMAX (https://maxlab.neuro.georgetown.edu/hmax.html) to extract the activations from the C2 layer of the model and compute Representational Dissimilarity Matrix (RDM) between 4 exemplars from the 10 categories used in the current study.

Category-Exemplar 1-Back Task

For the main experiment, participants completed a variation of a 1-back task, coined the "category-exemplar 1-back" in the 3T scanner. We created this new 1-back task to ensure that participants were attending carefully to each individual object, given our interest in fine-grained object discrimination. Like in a classic 1-back task, participants were shown a stream of individual objects and asked to indicate with a button press when the object was an exact repeat of the object previous to it, and no response was required when the object was from a different category as the one previous (Figure 2A). Our novel

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

twist was the addition of a second response option, whereby participants were asked to indicate with a different button when the object was from the same category as the previous one, but a different exemplar. The two response trial types served as catch trials to ensure participants' attention focused on differences between objects across consecutive trials, and to assess behavioural performance. These modifications of the classic 1-back task were introduced to ensure that participants attended closely to each individual object and engaged in object processing at the exemplar level. Specifically, successful identification of the repetition of different exemplars from the same category could not be based on local low-level features, such as changes in luminance, texture, or shape across consecutive trials. Participants used their right index and middle finger to respond, which was counterbalanced across participants. Of the three trial types—exemplar repeat, category repeat, different category—only the no response trials (i.e., different category) were used in the fMRI analysis to avoid motor confounds associated with button-presses. By extension, none of the trials considered for assessment of similarity in activation patterns were immediate neighbours.

Participants completed a total of eight functional runs that each lasted 4 min (stimulus duration = 1.2 s, inter-trial interval = 1 s). Run order was counter-balanced across participants. Within each run, each of the 40 exemplars were presented three times as no-response trials, and once as a catch trial, for a total of 24 presentations on no-response trials and 8 catch trials (excluded from fMRI analyses) per exemplar across the entire experiment. Prior to scanning, each participant completed a 5 min practice task with images from categories not included in the functional scanning experiment.

fMRI Data Acquisition

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

MRI data were acquired using a 3 T MR system (Siemens). A 32-channel head coil was used. Before the fMRI session, a whole head MP-RAGE volume (TE = 2.28 ms , TR = 2400 ms , TI = 1060 ms, resolution= 0.8 X 0.8 X 0.8 mm isometric) was acquired. This was followed by four fMRI runs, each with 300 volumes, which consisted of 42 T2*-weighted slices with a resolution of 1.7 X 1.7 mm (TE = 30 ms, TR = 1000 ms, slice thickness 1.7 mm , FOV 200 mm, parallel imaging with grappa factor 2). T2* weighted data were collected at this ultra-high resolution so as to optimize differentiation of BOLD signal in anterolateral versus posterior medial entorhinal cortex. The T2* slices were acquired in odd-even interleaved fashion in the anterior to posterior direction. Subsequently, a T2-weighted image (TE = 564 ms, TR = 3200 ms, resolution 0.8 X 0.8 X 0.8 mm isometric) was acquired. Finally, participants then completed four more fMRI runs. Total duration of MRI acquisition was approximately 60 min.

Preprocessing and Modeling

MRI data were converted to brain imaging data structure (BIDS; Gorgolewski et al., 2016) and ran through fmriprep-v1.1.8 (Esteban et al., 2018). This preprocessing included: motion correction, slice time correction, susceptibility distortion correction, registration from EPI to T1w image, and confounds estimated (e.g., tCompCor, aCompCor, framewise displacement). Component based noise correction was performed using anatomical and temporal CompCor, aCompCor and tCompCor, by adding these confound estimates as regressors in SPM12 during first level GLM (Behzadi, Restom, Liau, & Liu, 2007). Each participant was co-registered to the participant-specific T1w image by fmriprep. First-level analyses were conducted in native space for each participant with no spatial smoothing to preserve ultra-high resolution patterns of activity

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

for MVPA. Exemplar-specific multi-voxel activity patterns were estimated in 40 separate general linear models using the mean activity of the no-response trials across runs.

Region of Interest (ROI) Definitions for fMRI Analyses

Anatomical regions of interest were defined using multiple techniques. Automated segmentation was employed to delineate PrC, ErC, and PhC (ASHS; Wisse et al., 2016). We manually segmented each ERC obtained from ASHS into anterolateral and posteromedial entorhinal cortex following a protocol developed by Olsen and colleagues (2017), which is derived from a functional connectivity study (Maass et al., 2016). A probabilistic atlas was used to define EVC (Wang et al., 2015) and TP (Fischl, 2012). A functional localizer was used to define LOC as the contiguous voxels located along the lateral extent of the occipital lobe that responded more strongly to intact objects than scrambled objects ($p<.01$, uncorrected; Proklova et al., 2016).

In the VVS, we focused on lateral occipital complex and the temporal pole as they have previously been linked to object processing (e.g., Grill-Spector, Kourtzi, & Kanwisher, 2001; Martin et al., 2018), as well as early visual cortex. In the MTL, we included ROIs for the posteromedial ErC and parahippocampal cortex, both of which have been linked to scene processing (e.g., Maass et al., 2015; Schroder et al., 2015; Schultz et al., 2015; Epstein & Baker, 2020).

Representational Similarity Analyses of fMRI data

For each ROI, linear correlation distances (Pearson's *r*) were calculated between all pairs of exemplar-specific multi-voxel patterns using CoSMoMVPA toolbox in Matlab (Oosterhof, Connolly, Haxby, 2016) across all voxels. These correlations were used to

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

create participant-specific brain-based RDMs (1-Pearson's r), which captures the unique neural pattern dissimilarities between all exemplars within each category (n=10), within each region (n=8).

Whole-volume RSA was conducted using surface-based searchlight analysis (Kriegeskorte et al., 2006; Oosterhof et al., 2011; Martin et al., 2018). Specifically, we defined a 100-voxel neighborhood around each surface voxel, and computed a brain-based RDM within this region, analogous to the ROI-based RSA. This searchlight was swept across the entire cortical surface (Kriegeskorte et al., 2006; Oosterhof et al., 2011). First, the entire perceived similarity RDM for all within category ratings was compared to each searchlight. These brain-behaviour correlations were Fisher transformed and mapped to the centre of each searchlight for each participant separately. Participant-specific similarity maps were then standardized and group-level statistical analysis was performed. Threshold-free cluster enhancement (TFCE) was used to correct for multiple comparisons with a cluster threshold of $p<.05$ (Smith and Nichols, 2009).

Acknowledgments

This work was supported by a Canadian Institutes for Health Research Project Grant (CIHR Grant # 366062) to A.K. and S.K.. K.F. was funded through a Natural Sciences and Engineering Research Council doctoral Canadian Graduate Scholarship (NSERC CGS-D) and an Ontario Graduate Scholarship (OGS). A.B was funded through an Ontario Trillium Scholarship for Doctoral study in Canada.

Author Contributions

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Kayla M. Ferko, Conceptualization, Software, Data curation, Formal analysis, Investigation, Visualization, Methodology, Writing—original draft, Project administration, Writing—review and editing; Anna Blumenthal, Conceptualization, Investigation, Data curation, Methodology, Writing—original draft, Writing—review and editing; Chris B. Martin, Formal Analysis, Writing—review and editing; Daria Proklova, Methodology, Sotware, Data curation, Formal analysis; Alexander N. Minos, Investigation, Data curation; Lisa M. Saksida, Conceptualization, Writing—review and editing; Timothy J. Bussey, Conceptualization, Writing—review and editing; Ali Khan, Supervision, Funding acquisition, Software, Data curation, Methodology; Stefan Köhler, Conceptualization, Supervision, Funding acquisition, Methodology, Formal, analysis, Project administration, Writing—original draft, Writing—review and editing.

Competing Interests

The authors declare no competing interests.

References

Banks, P. J., Warburton, E. C., Brown, M. W., & Bashir, Z. I. (2014). Mechanisms of synaptic plasticity and recognition memory in the perirhinal cortex. Progress in molecular biology and translational science, 122, 193-209. DOI: https://doi.org/10.1016/B978-0-12-420170-5.00007-6

Barense, M. D., Gaffan, D., & Graham, K. S. (2007). The human medial temporal lobe processes online representations of complex objects. Neuropsychologia, 45(13), 2963-2974. DOI: https://doi.org/10.1016/j.neuropsychologia.2007.05.023

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Barense, M. D., Henson, R. N., Lee, A. C., & Graham, K. S. (2010). Medial temporal lobe activity during complex discrimination of faces, objects, and scenes: Effects of viewpoint. Hippocampus, 20(3), 389-401. DOI: https://doi.org/10.1002/hipo.20641

Bartko, S. J., Winters, B. D., Cowell, R. A., Saksida, L. M., & Bussey, T. J. (2007). Perceptual functions of perirhinal cortex in rats: zero-delay object recognition and simultaneous oddity discriminations. Journal of Neuroscience, 27(10), 2548-2559. DOI: https://doi.org/10.1523/JNEUROSCI.5171-06.2007

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage, 37(1), 90-101. DOI: https://doi.org/10.1016/j.neuroimage.2007.04.042

Bonnen, T., Yamins, D. L., & Wagner, A. D. (2021). When the ventral visual stream is not enough: A deep learning account of medial temporal lobe involvement in perception. Neuron, 109(17), 2755-2766. DOI: https://doi.org/10.1016/j.neuron.2021.06.018

Bruffaerts, R., Dupont, P., Peeters, R., De Deyne, S., Storms, G., & Vandenberghe, R. (2013). Similarity of fMRI activity patterns in left perirhinal cortex reflects semantic similarity between words. Journal of Neuroscience, 33(47), 18597-18607. DOI: https://doi.org/10.1523/JNEUROSCI.1548-13.2013

Buckley, M. J., Booth, M. C., Rolls, E. T., & Gaffan, D. (2001). Selective perceptual impairments after perirhinal cortex ablation. Journal of Neuroscience, 21(24), 9824-9836. DOI: https://doi.org/10.1523/JNEUROSCI.21-24-09824.2001

Buckley, M. J., & Gaffan, D. (2006). Extent of MTL lesions in animals and human patients. Trends in Cognitive Sciences, 3(10), 100-107.

Bussey, T. J., & Saksida, L. M. (2007). Memory, perception, and the ventral visual-perirhinal-hippocampal stream: thinking outside of the boxes. Hippocampus, 17(9), 898-908.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Bussey, T. J., Saksida, L. M., & Murray, E. A. (2002). Perirhinal cortex resolves feature ambiguity in complex visual discriminations. European Journal of Neuroscience, 15(2), 365-374.

Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M., & Poggio, T. (2007). A model of V4 shape selectivity and invariance. Journal of neurophysiology, 98(3), 1733-1750.

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. Proceedings of the National Academy of Sciences, 111(40), 14565-14570. DOI: https://doi.org/10.1073/pnas.1402594111

Charest, I., & Kriegeskorte, N. (2015). The brain of the beholder: honouring individual representational idiosyncrasies. Language, Cognition and Neuroscience, 30(4), 367-379. DOI: https://doi.org/10.1080/23273798.2014.1002505

Cichy, R. M., Kriegeskorte, N., Jozwik, K. M., van den Bosch, J. J., & Charest, I. (2019). The spatiotemporal neural dynamics underlying perceived similarity for real-world objects. Neuroimage, 194, 12-24.

Clarke, A., & Tyler, L. K. (2014). Object-specific semantic coding in human perirhinal cortex. Journal of Neuroscience, 34(14), 4766-4775. DOI: https://doi.org/10.1523/JNEUROSCI.2828-13.2014

Collins, E., & Behrmann, M. (2020). Exemplar learning reveals the representational origins of expert category perception. Proceedings of the National Academy of Sciences, 117(20), 11167-11177. DOI: https://doi.org/10.1073/pnas.1912734117

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., ... & Haxby, J. V. (2012). The representation of biological classes in the human brain. Journal of Neuroscience, 32(8), 2608-2618. DOI: https://doi.org/10.1523/JNEUROSCI.5547-11.2012

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Connor, C. E., & Knierim, J. J. (2017). Integration of objects and space in perception and memory. Nature neuroscience, 20(11), 1493. DOI: https://doi.org/10.1038/nn.4657

Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2010). Components of recognition memory: dissociable cognitive processes or just differences in representational complexity?. Hippocampus, 20(11), 1245-1262.

Epstein, R. A., & Baker, C. I. (2019). Scene perception in the human brain. DOI: 10.1146/annurev-vision-091718-014809

Erez, J., Cusack, R., Kendall, W., & Barense, M. D. (2016). Conjunctive coding of complex object features. Cerebral Cortex, 26(5), 2271-2282. DOI: https://doi.org/10.1093/cercor/bhv081

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., ... & Oya, H. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature methods, 16(1), 111-116. DOI: https://doi.org/10.1038/s41592-018-0235-4

Fischl, B. (2012). FreeSurfer. Neuroimage, 62(2), 774-781. DOI: https://doi.org/10.1016/j.neuroimage.2012.01.021

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. Cognition, 52(2), 125-157. DOI: https://doi.org/10.1016/0010-0277(94)90065-5

Goldstone, R. L., & Son, J. Y. (2012). Similarity. Oxford University Press.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., ... & Handwerker, D. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Scientific data, 3(1), 1-9. DOI: https://doi.org/10.1038/sdata.2016.44

Graham, K. S., Barense, M. D., & Lee, A. C. (2010). Going beyond LTM in the MTL: a synthesis of neuropsychological and neuroimaging findings on the role of the medial temporal lobe

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

in memory and perception. Neuropsychologia, 48(4), 831-853. DOI: https://doi.org/10.1016/j.neuropsychologia.2010.01.001

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. Vision research, 41(10-11), 1409-1422. DOI: https://doi.org/10.1016/S0042-6989(01)00073-6

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. Nature Reviews Neuroscience, 15(8), 536-548.

Hales, J. B., & Clark, R. E. (2015). Organization and structure of brain systems supporting memory. In The Maze Book (pp. 143-164). Humana Press, New York, NY. DOI: https://doi.org/10.1007/978-1-4939-2159-1_6

Haushofer, J., Livingstone, M. S., & Kanwisher, N. (2008). Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. PLoS Biol, 6(7), e187. DOI: https://doi.org/10.1371/journal.pbio.0060187

Hebart, M., Zheng, C. Y., Pereira, F., & Baker, C. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgments.

Inhoff, M. C., Heusser, A. C., Tambini, A., Martin, C. B., O'Neil, E. B., Köhler, S., ... & Davachi, L. (2019). Understanding perirhinal contributions to perception and memory: Evidence through the lens of selective perirhinal damage. Neuropsychologia, 124, 9-18. DOI: https://doi.org/10.1016/j.neuropsychologia.2018.12.020

Iordan, M. C., Ellis, C. T., Lesnick, M., Osherson, D. N., & Cohen, J. D. (2018). Feature Ratings and Empirical Dimension-Specific Similarity Explain Distinct Aspects of Semantic Similarity Judgments. In CogSci.

Kent, B. A., Hvoslef-Eide, M., Saksida, L. M., & Bussey, T. J. (2016). The representational–hierarchical view of pattern separation: Not just hippocampus, not just space, not just

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

memory?. Neurobiology of learning and memory, 129, 99-106. DOI: https://doi.org/10.1016/j.nlm.2016.01.006

Kubilius, J., Bracci, S., & Op de Beeck, H. P. (2016). Deep neural networks as a computational model for human shape sensitivity. *PLoS computational biology*, *12*(4), e1004896. https://doi.org/10.1371/journal.pcbi.1004896

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS computational biology, 10(11), e1003915.

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. Trends in cognitive sciences, 17(8), 401-412. DOI: https://doi.org/10.1016/j.tics.2013.06.007

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. Frontiers in psychology, 3, 245. DOI: https://doi.org/10.3389/fpsyg.2012.00245

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron, 60(6), 1126-1141. DOI: https://doi.org/10.1016/j.neuron.2008.10.043

Levy, D. A., Shrager, Y., & Squire, L. R. (2005). Intact visual discrimination of complex and feature-ambiguous stimuli in the absence of perirhinal cortex. Learning & memory, 12(1), 61-66. DOI: 10.1101/lm.84405

Lee, A. C., Brodersen, K. H., & Rudebeck, S. R. (2013). Disentangling spatial perception and spatial memory in the hippocampus: a univariate and multivariate pattern analysis fMRI study. Journal of cognitive neuroscience, 25(4), 534-546. DOI: https://doi.org/10.1162/jocn_a_00301

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Lee, A. C., Scahill, V. L., & Graham, K. S. (2008). Activating the medial temporal lobe during oddity judgment for faces and scenes. Cerebral Cortex, 18(3), 683-696. DOI: https://doi.org/10.1093/cercor/bhm104

Liang, J. C., Erez, J., Zhang, F., Cusack, R., & Barense, M. D. (2020). Experience Transforms Conjunctive Object Representations: Neural Evidence for Unitization After Visual Expertise. Cerebral Cortex, 30(5), 2721-2739. DOI:https://doi.org/10.1093/cercor/bhz250

Maass, A., Berron, D., Libby, L. A., Ranganath, C., & Düzel, E. (2015). Functional subregions of the human entorhinal cortex. Elife, 4, e06426. DOI: https://doi.org/10.7554/eLife.06426.001

Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L., & Barense, M. D. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. Elife, 7, e31873. DOI: https://doi.org/10.7554/eLife.31873.001

Martin, C. B., McLean, D. A., O'Neil, E. B., & Köhler, S. (2013). Distinct familiarity-based response patterns for faces and buildings in perirhinal and parahippocampal cortex. *Journal of Neuroscience*, *33*(26), 10915-10923. DOI: https://doi.org/10.1523/JNEUROSCI.0126-13.2013

Martin, C. B., Sullivan, J. A., Wright, J., & Köhler, S. (2018). How landmark suitability shapes recognition memory signals for objects in the medial temporal lobes. *NeuroImage*, *166*, 425-436. DOI: https://doi.org/10.1016/j.neuroimage.2017.11.004

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. Psychological review, 100(2), 254. DOI: https://doi.org/10.1037/0033-295X.100.2.254

Migo, E. M., Montaldi, D., & Mayes, A. R. (2013). A visual object stimulus database with standardized similarity information. Behavior research methods, 45(2), 344-354. DOI: https://doi.org/10.3758/s13428-012-0255-4

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Minos, A. N., Ferko, K. M., & Köhler, S. (2021). Expertise predicts perceived visual similarity between exemplars of real-world object categories. Journal of Vision, 21(9), 2985-2985. DOI: https://doi.org/10.1167/jov.21.9.2985

Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. Neuroimage, 53(1), 103-118. DOI: https://doi.org/10.1016/j.neuroimage.2010.05.051

Moutsiana, C., De Haas, B., Papageorgiou, A., Van Dijk, J. A., Balraj, A., Greenwood, J. A., & Schwarzkopf, D. S. (2016). Cortical idiosyncrasies predict the perception of object size. Nature communications, 7(1), 1-12. DOI: https://doi.org/10.1038/ncomms12110

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. Social cognitive and affective neuroscience, 4(1), 101-109. DOI: https://doi.org/10.1093/scan/nsn044

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. Frontiers in psychology, 4, 128. DOI: https://doi.org/10.3389/fpsyg.2013.00128

Murray, E. A., & Bussey, T. J. (1999). Perceptual–mnemonic functions of the perirhinal cortex. Trends in cognitive sciences, 3(4), 142-151. DOI: https://doi.org/10.1016/S1364-6613(99)01303-0

O'Neil, E. B., Barkley, V. A., & Köhler, S. (2013). Representational demands modulate involvement of perirhinal cortex in face processing. Hippocampus, 23(7), 592-605. DOI: https://doi.org/10.1002/hipo.22117

O'Neil, E. B., Cate, A. D., & Köhler, S. (2009). Perirhinal cortex contributes to accuracy in recognition memory and perceptual discriminations. *Journal of Neuroscience*, *29*(26), 8329-8334. DOI: https://doi.org/10.1523/JNEUROSCI.0374-09.2009

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. Frontiers in neuroinformatics, 10, 27. DOI: https://doi.org/10.3389/fninf.2016.00027

Op de Beeck, H. P. O., Torfs, K., & Wagemans, J. (2008). Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. Journal of Neuroscience, 28(40), 10111-10123. DOI: https://doi.org/10.1523/JNEUROSCI.2511-08.2008

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling representations of object shape and object category in human visual cortex: The animate–inanimate distinction. Journal of cognitive neuroscience, 28(5), 680-692. DOI: https://doi.org/10.1162/jocn_a_00924

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature neuroscience, 2(11), 1019-1025.

Schröder, T. N., Haak, K. V., Jimenez, N. I. Z., Beckmann, C. F., & Doeller, C. F. (2015). Functional topography of the human entorhinal cortex. Elife, 4, e06738. DOI: https://doi.org/10.7554/eLife.06738.001

Schultz, H., Sommer, T., & Peters, J. (2015). The role of the human entorhinal cortex in a representational account of memory. Frontiers in Human Neuroscience, 9, 628. DOI: https://doi.org/10.3389/fnhum.2015.00628

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. Nature neuroscience, 14(1), 28-30. DOI: https://doi.org/10.1038/nn.2706

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. Proceedings of the national academy of sciences, 104(15), 6424-6429. Categorization

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage, 44(1), 83-98. DOI: https://doi.org/10.1016/j.neuroimage.2008.03.061

Stark, C. E., & Squire, L. R. (2000). Intact visual perceptual discrimination in humans in the absence of perirhinal cortex. Learning & Memory, 7(5), 273-278. DOI: 10.1101/lm.35000

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. Cerebral cortex, 25(10), 3911-3931. DOI: https://doi.org/10.1093/cercor/bhu277

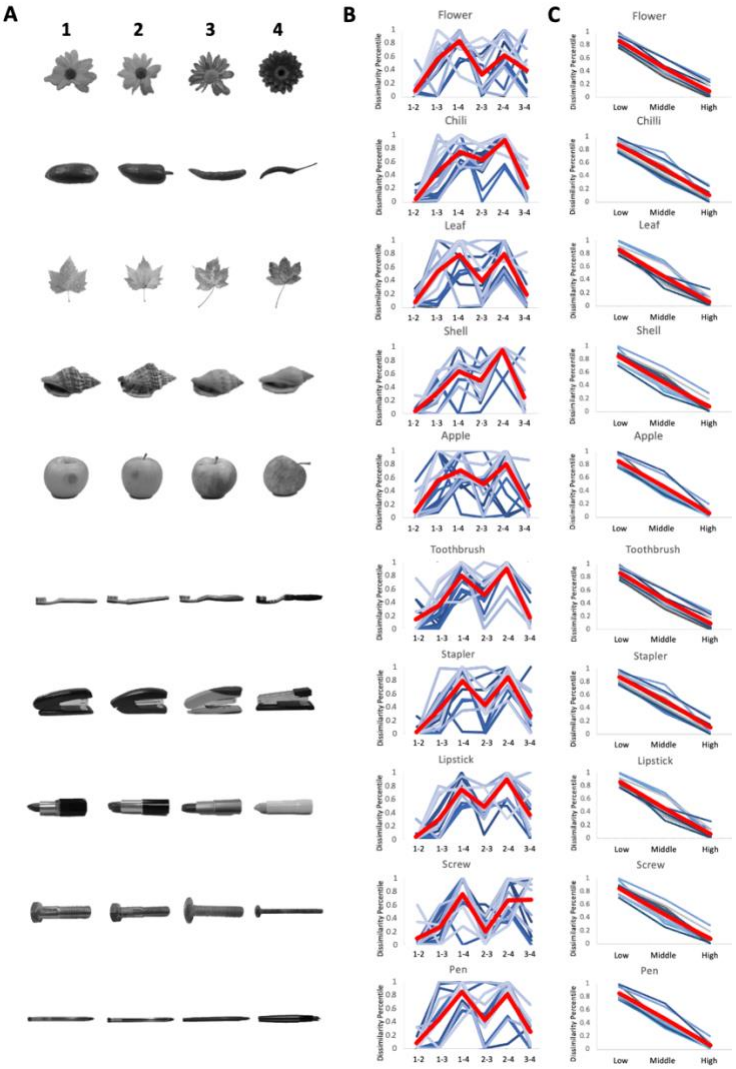
Wisse, L. E., Kuijf, H. J., Honingh, A. M., Wang, H., Pluta, J. B., Das, S. R., ... & Geerlings, M. I. (2016). Automated hippocampal subfield segmentation at 7T MRI. American Journal of Neuroradiology, 37(6), 1050-1057. DOI: https://doi.org/10.3174/ajnr.A4659

Yeung, L. K., Olsen, R. K., Bild-Enkin, H. E., D'Angelo, M. C., Kacollja, A., McQuiggan, D. A., ... & Barense, M. D. (2017). Anterolateral entorhinal cortex volume predicted by altered intra-item configural processing. Journal of Neuroscience, 37(22), 5527-5538. DOI: https://doi.org/10.1523/JNEUROSCI.3664-16.2017

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

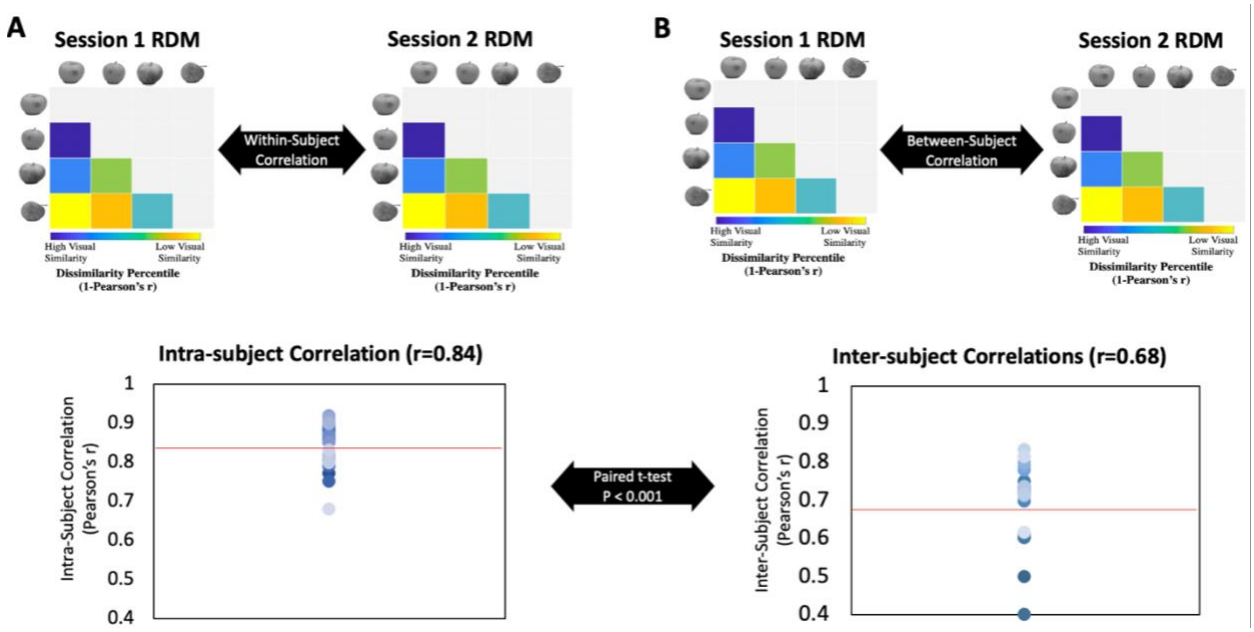
Supplementary Materials

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



Supplementary Figure 1. (A) 40 images of object stimuli sampled from Migo et al. (2013) and employed in current study. Images of objects are from 10 categories with 4 exemplars in each category. Dimensions on which exemplars differ include: external shape envelope, texture, and internal spatial configuration of parts. **(B)** Behavioural results from iMDS task obtained off-line from all 24 participants in the fMRI study (Mean in thick red line). Perceived similarity ratings (expressed as Dissimilarity Percentile) were obtained separately for each category. They are displayed for all pairwise comparisons (with numbers indicating specific exemplars as shown in A) in arbitrary order along x-axis. Ratings reveal some variability across participants in all categories, which is also reflected in intersubject correlations shown in Fig. 2c in Main text. **(C)** Ratings from (B) reordered according to participant's own rank ordering of perceived similarity. We grouped the 2 lowest, 2 middle, and 2 highest similarities in observer-specific manner. Shown dissimilarity percentiles reflect average values for 2 pairs of ratings in each grouped ranking. Notably, the range of dissimilarity percentile values was comparable across the three levels ($p > 0.5$). Displayed rank-ordered data were employed to assess sensitivity of behaviour and fMRI responses to participants' own perceived similarity of objects.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



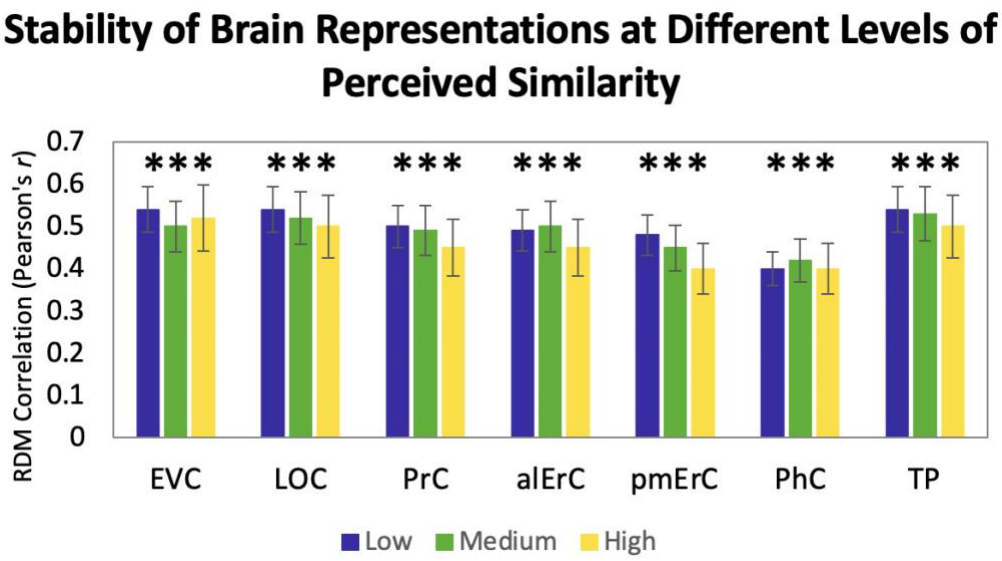
Supplementary Figure 2. In a follow-up study, a distinct group of 30 participants completed two sessions of the iMDS task for the 10 object categories separated by 7 days +/- 1 day later. **(A)** Intra-subject correlations for perceived similarity ratings across all exemplars and categories. Correlations were computed between each participants' perceived similarity Representational Dissimilarity Matrices (RDMs) from Session 1 and from Session 2. The mean within-subject correlation across the two sessions was 0.84, indicating high stability of participant's perceived similarity ratings one week apart. **(B)** Intersubject correlations for perceived similarity ratings across all exemplars and categories. Correlations were computed between each participant's RDM Session 1 with the mean RDM (excluding the participant) in Session 2. Mean inter-subject correlation was 0.68. Critically, a paired t-test(intra-subject>inter-subject correlation) confirmed that intra-subject correlations were significantly higher than inter-subject correlations ($p<0.0001$). This pattern of results indicates that the perceived similarity structure that is unique to the individual observer is a stable characteristic.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

				MEAN PROPORTION OF RESPONSES (SD)		
				"SAME EXEMPLAR"	"SAME CATEGORY"	NO RESPONSE
TRIAL TYPE	EXEMPLAR REPEAT			**.91 (.032)**	.07 (.01)	.02 (.001)
	CATEGORY REPEAT	Level of Perceived Similarity	High	.36 (.019)	**.59 (.029)**	.05 (.01)
			Middle	.18 (.017)	**.79 (.025)**	.03 (.007)
			Low	.07 (.01)	**.90 (.03)**	.03 (.006)
	DIFFERENT CATEGORY			.01 (.009)	.06 (.014)	**.93 (.02)**

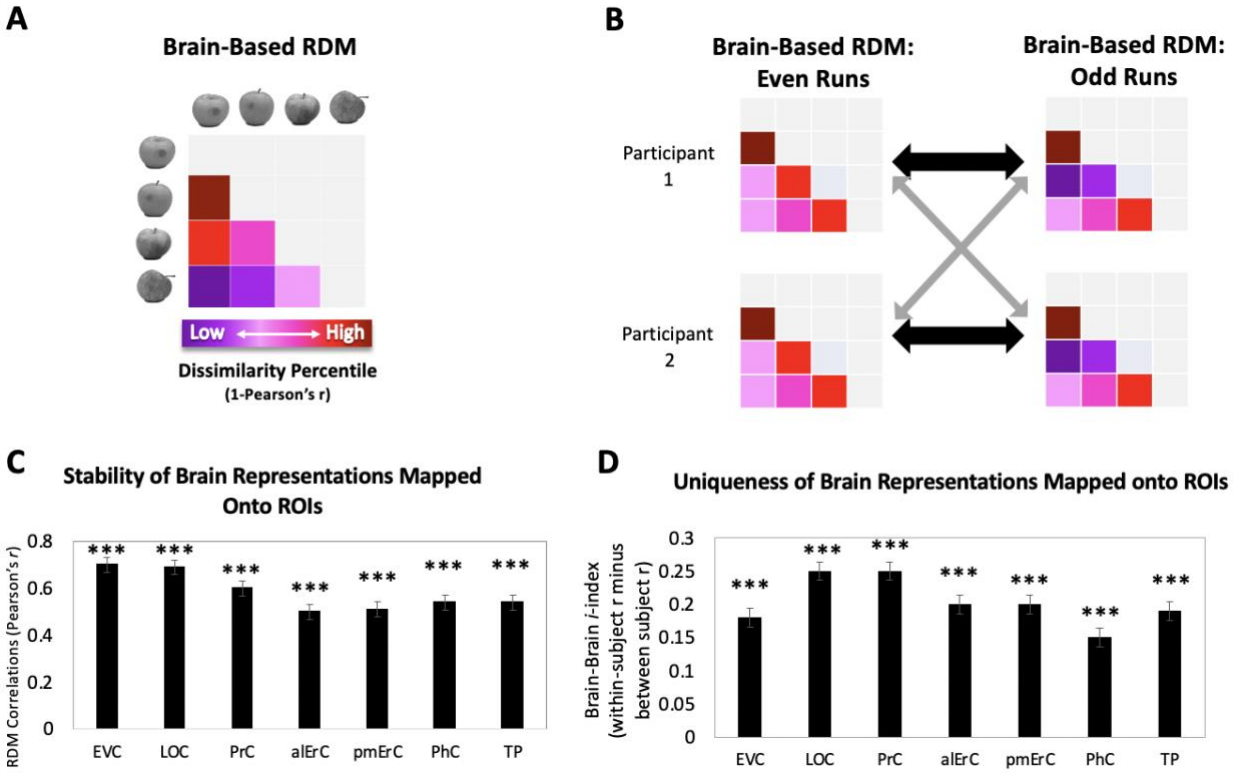
Supplementary Table 1. Behavioural performance on Category-Exemplar 1-Back Task. Each object image belonged to one of three trial types: exemplar repeat (Same Exemplar), category repeat (Same Category, different exemplar), no repeat (Different Category, different exemplar). Proportion of correct responses for each trial type are indicated in green. Same Category trial types are further divided according to level of perceived visual similarity, measured using participants ratings on the iMDS object sorting task. Behavioural performance reveals sensitivity to perceived similarity in terms of accuracy of responses.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



Supplementary Figure 3. Relationship between Brain-based RDMs in even and odd runs, within and between participants at different levels of perceived similarity. Correlation of brain-based RDMs for even and odd runs at low, medium, and high levels of similarity. Activation patterns in all ROIs show significant correlations at all levels of perceived similarity (***P<.001, Bonferroni-corrected for regions and levels).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



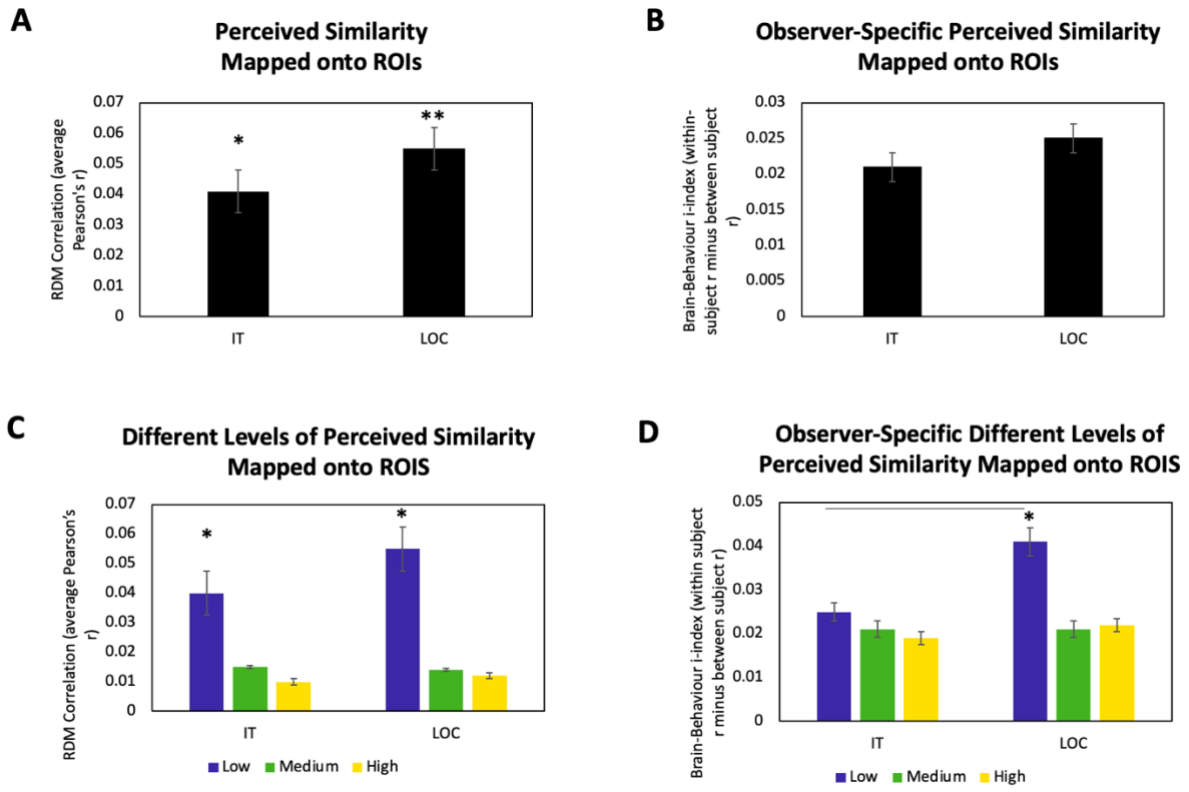
Supplementary Figure 4. Relationship between Brain-based RDMs in even and odd runs, within and between participants. **(A)** Pairwise pattern dissimilarity percentiles were used to create observer-specific brain-based RDMs for even and odd runs. **(B)** Brain-based RDMs from even runs were correlated with (i) observers' brain-based RDMs for odd runs (black double arrows = within subject r) and (ii) other observers' brain-based RDMs from odd runs (grey arrows = between subject r) for comparison. **(C)** Patterns in all ROIs show stable similarity structure as reflected in significant correlations between odd and even run; ***P<.001, Bonferroni-corrected based on regions; Error bars represent SEM. **(D)** Patterns in all ROIs are unique to each individual to each individual as reflected in Brain-Brain *i*-index (i.e., within minus between subject correlation; ***P<.001, Bonferroni-corrected based on regions, with testing against a null-distribution created by randomizing subject-labels; Error bars represent SEM estimated based on randomization). EVC = Early Visual Cortex; LOC = Lateral Occipital Complex; PrC = Perirhinal Cortex; alErC = anterolateral entorhinal cortex; pmErC = posteromedial entorhinal cortex; PhC = Parahippocampal cortex; TP = Temporal Pole).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

	Beta (own)		**Beta (average)**		**Beta (own) – Beta (average)**	
	T-value	**P-value**	**T-value**	**P-value**	**T-value**	**P-value**
EVC	4.56	**0.007***	5.12	**0.0057***	0.65	0.600
LOC	4.99	**0.0105***	7.43	**0.00098***	1.55	0.10
PrC	7.01	**0.0024***	3.06	**0.0686**	4.83	**0.008***
alErC	6.50	**0.0202***	3.01	**0.0699**	4.59	**0.007***
pmErC	0.98	0.59	1.53	0.23	1.09	0.39
PhC	1.13	0.42	1.13	0.37	1.70	0.25
TP	1.74	0.12	1.74	0.22	0.95	0.50

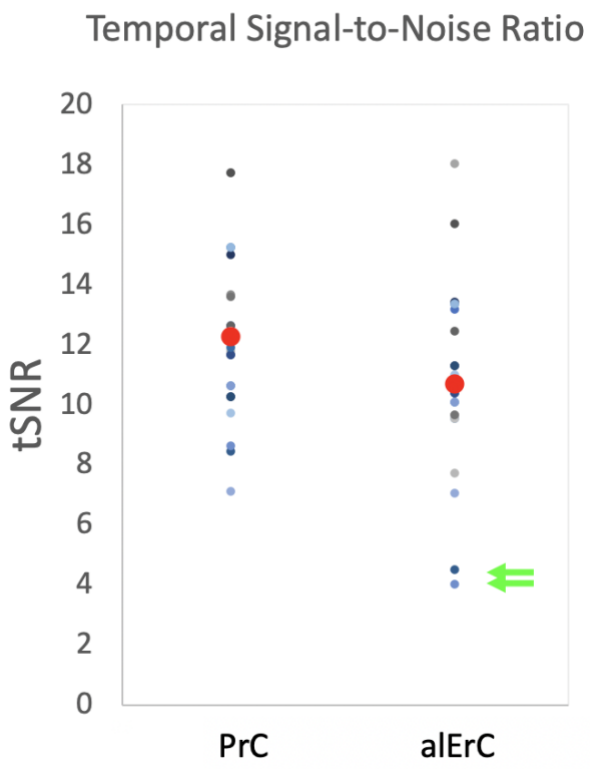
Supplementary Table 2. Multiple linear regression: Brain RDM ~ (own RDM + average RDM). To determine the contributions of shared perceived similarity versus observer-specific perceived similarity to activity patterns in our regions of interest (ROIs), we conducted multiple regression analyses for the relationships depicted in Figure 3D using average similarity ratings and observer's own ratings as predictors. Beta-values were tested against 0 in ROI using paired t-tests with Bonferroni-correction. Beta(own) was significantly higher than 0 in EVC, LOC, PRC, and alErC. Beta(average) was significant in EVC and LOC, but not PRC, alErC, nor in the other ROIs. Critically, the difference between Beta(own) and Beta(average) was significant only in PrC and alErC. The latter pattern of results converges with the results obtained based on calculation of the i-index as shown in Figure 7B in the main text.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



Supplementary Figure 5. Comparison between Inferotemporal (IT) Cortex and Lateral Occipital Complex (LOC): Brain-based representational dissimilarity matrices (RDMs) and their relationship to perceived visual similarity **(A)** at all levels of perceived similarity and **(B)** corresponding i-index; and **(C)** at different levels of perceived similarity and **(D)** corresponding i-index. These 4 analyses from the main experiment were included here to compare results in IT cortex with those in LOC. IT cortex is a large piece of cortex that extends across occipital and temporal cortex. This region of interest (similar to the one used in Charest et al., 2014) overlaps with regions of interest used in the main experiment including: LOC, PhC and TP. We were specifically interested in LOC—the object selective portion of larger IT cortex—because our stimuli were objects. The results across these four sets of analyses were highly similar between the two regions, with the only noticeable exception being LOC activity patterns predicting observer-specific perceived similarity at low levels of similarity while IT patterns not predicting any observer-specific similarity at all.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.21.427602; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



Supplementary Figure 6. Temporal Signal-to-Noise Ratio in PrC (M=12.25, SD=2.68) and alErC (M=10.80, SD=3.05) in each of the 25 participants (average denoted by bolded red dot). Green arrows indicate 2 participants with tSNR in alERC below 2SD of the Mean. These participants were excluded from fMRI analyses