Visual homogeneity computations in the brain enable solving generic visual tasks

Georgin Jacob[1,2], R. T. Pramod[1,2], and S. P. Arun[2,1]*

[1]Department of Electrical Communication Engineering & [2]Centre for Neuroscience

Indian Institute of Science, Bangalore 560012

*Correspondence to sparun@iisc.ac.in

ABSTRACT

Many visual tasks involve looking for specific object features. But we also often perform generic tasks where we look for specific property, such as finding an odd item, deciding if two items are same, or if an object has symmetry. How do we solve such tasks? Using well-known principles governing multiple object representations, we show that displays with repeating elements can be distinguished from heterogeneous displays using a property we define as visual homogeneity. In behavior, visual homogeneity predicted response times on visual search and symmetry tasks. Brain imaging during these tasks revealed that visual homogeneity in both tasks is localized to a region in the object-selective cortex. Thus, a novel image property, visual homogeneity, is encoded in a localized brain region, to solve generic visual tasks.

SIGNIFICANCE STATEMENT

Most visual tasks involve looking for specific features, like finding a face in a crowd. But we often also perform generic tasks where we look for a particular image property – such as finding an odd item, deciding if two items are same, or judging if an object is symmetric. Precisely how we solve such disparate tasks is unclear. Here, we show that these tasks can be solved using a simple property we define as visual homogeneity. In behavior, visual homogeneity predicted response times on visual search and symmetry tasks. In brain imaging, it was localized to a region near the object-selective cortex. Thus, a novel image property, visual homogeneity, is computed by the brain to solve generic visual tasks.

INTRODUCTION

Many visual tasks involve looking for specific objects or features, such as a friend in a crowd or selecting vegetables in the market. In such tasks, which have been studied extensively, we form a template in our brain that helps guide eye movements and locate the target (Peelen and Kastner, 2014). This template becomes a decision variable that can be used to solve the task: the degree of match to the template indicates the presence or absence of the desired object. However, we also easily perform tasks that do not involve any specific feature template but rather involve finding a property or relation between items. Examples of these generic tasks include finding an odd item, deciding if two items are same and judging if an object is symmetric. While machine vision algorithms are extremely successful in solving feature-based tasks like object categorization (Serre, 2019), they find generic same-different tasks difficult to solve (Kim et al., 2018; Ricci et al., 2021; Puebla and Bowers, 2022), although recent studies have found certain architectures to generalize quite well (Tartaglini et al., 2023). How do we perform such tasks with equal ease? What is the underlying template or decision variable that we use to solve generic visual tasks?

To start with, these tasks appear completely different at least in the way we describe them verbally. Even theoretical studies have considered visual search (Verghese, 2001; Wolfe and Horowitz, 2017), same-different judgments (Nickerson, 1969; Petrov, 2009) and symmetry detection (Wagemans, 1997; Bertamini and Makin, 2014) as conceptually separate tasks. However, at a deeper level, these tasks are similar because they all involve discriminating between items with repeating features from those without repeating features. We reasoned that if images with repeating

features are somehow represented differently in the brain, this difference could be used to solve all these tasks without requiring separate computations for each task. Here we provide evidence for this hypothesis through behavioural and brain imaging experiments on humans.

Our key predictions are depicted in Figure 1. Consider a visual search task where participants have to indicate if a display contains an oddball target (Figure 1A) or not (Figure 1B). According to the well-known principle of divisive normalization in high-level visual cortex (Zoccolan et al., 2005; Agrawal et al., 2020; Katti and Arun, 2022), the neural response to multiple objects is the average of the single object responses. Accordingly, the response to an array of identical items will be the same as the response to the single item. Moreover, the response to an array containing a target among distractors would lie along the line joining the target and distractor in the (neural) representational space. These possibilities are shown for all possible arrays made from three objects in Figure 1C. It can be seen that the homogeneous (target-absent) arrays stand apart since they contain repeating items, whereas the heterogeneous (target-present) arrays come closer since they contain a mixture of items. Since displays with repeating items are further away from the center of this space, this distance can be used to discriminate them from heterogeneous displays (Figure 1C, *inset*). We note here that this decision process does not capture many other complex forms of visual search but may guide the initial selection process towards possible targets.

We reasoned similarly for symmetry detection: here, participants have to decide if an object is asymmetric (Figure 1D) or symmetric (Figure 1E). According to multiple object normalization, objects with two different parts would lie along the line joining

objects containing the two repeated parts (Figure 1F). Indeed, both symmetric and asymmetric objects show part summation in their neural responses (Pramod and Arun, 2018). Consequently, symmetric objects will be further away from the centre of this space compared to asymmetric objects, and this can be the basis for discriminating them (Figure 1F, *inset*).

We define this distance from the center for each image as its *visual homogeneity* (VH). We made two key experimental predictions to test in behavioural and brain imaging experiments. First, if VH is being used to solve visual search and symmetry detection tasks, then responses should be slowest for displays with VH close to the decision boundary and faster for displays with VH far away (Figure 1G). This predicts opposite correlations between response time and VH: for target-present arrays and asymmetric objects, the response time should be positively correlated with VH. By contrast, for target-absent arrays and symmetric objects, response time should be negatively correlated with VH. Importantly, because response times can be positively or negatively correlated with VH, the net correlation between response time and VH will be close to zero. Second, if VH is encoded by a dedicated brain region, then brain activity in that region will be positively correlated with VH (Figure 1H). Such a positive correlation cannot be explained easily by cognitive processes linked to response time such as attention or task difficulty, since response times will have zero correlation with the mean activity of this region.

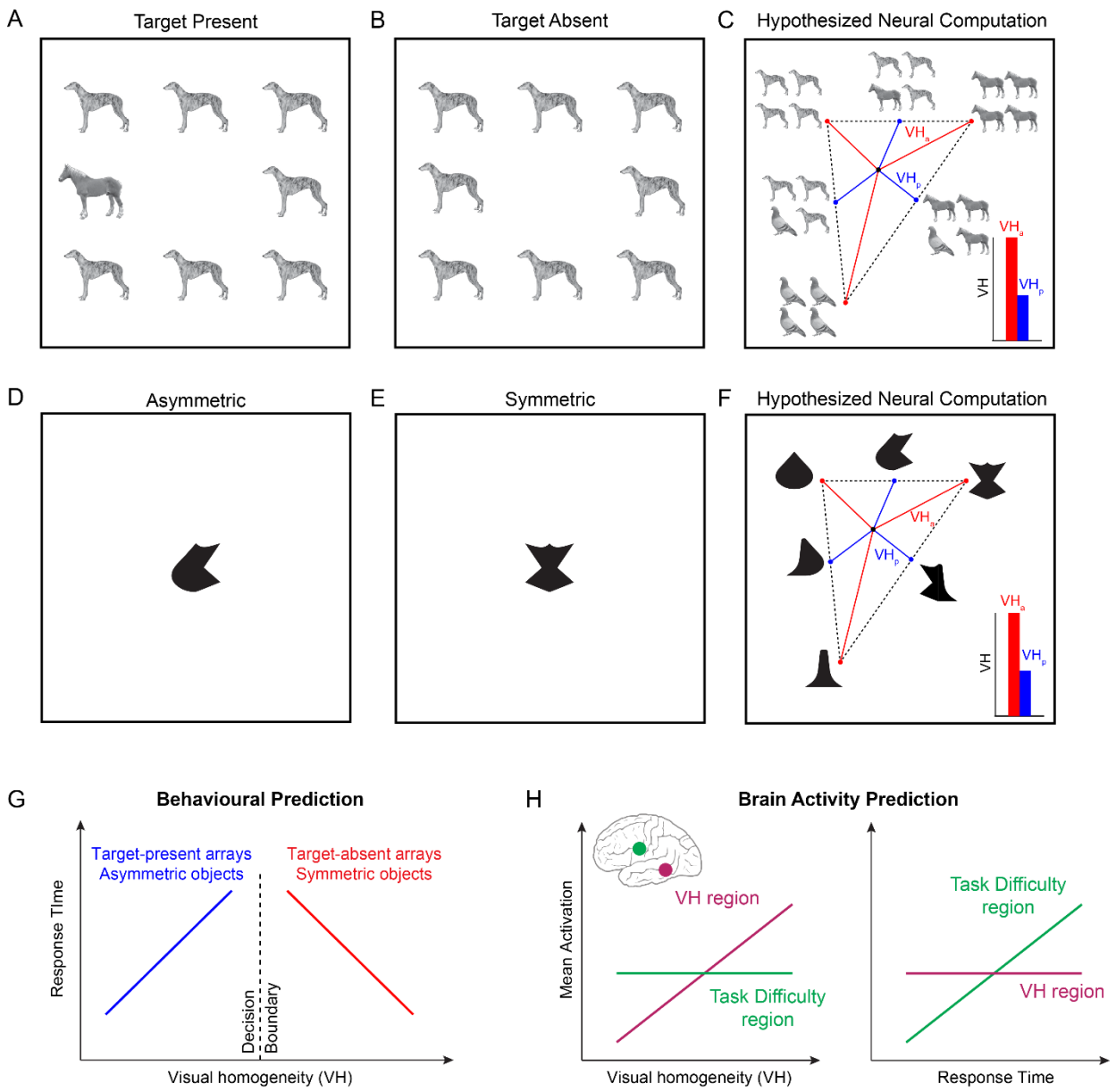

Figure 1. Solving visual search and symmetry tasks using visual homogeneity.

(A) Example target-present search display, containing a single oddball target (horse) among identical distractors (dog). Participants in such tasks have to indicate whether the display contains an oddball or not, without knowing the features of the target or distractor. This means they have to perform this task by detecting some property of each display rather than some feature contained in it.

(B) Example target-absent search display containing no oddball target.

(C) Hypothesized neural computation for target present/absent judgements. According to multiple object normalization, the response to multiple items is an average of the responses to the individual items. Thus, the response to a target-absent array will be identical to the individual items, whereas the response to a target-present array will lie along the line joining the corresponding target-absent arrays. This causes the target-absent arrays to stay apart (*red lines*), and the target-present arrays to come closer due to mixing (*blue lines*). If we calculate

the distance (VH, for visual homogeneity) for each display, then target-absent arrays will have a larger distance to the center (VH_a) compared to target-present arrays (VH_p), and this distance can be used to distinguish between them. *Inset:* Schematic distance from center for target-absent arrays (red) and target-present arrays (blue). Note that this approach might only reflect the initial target selection process involved in oddball visual search but does not capture all forms of visual search.

(D) Example asymmetric object in a symmetry detection task. Here too, participants have to indicate if the display contains a symmetric object or not, without knowing the features of the object itself. This means they have to perform this task by detecting some property in the display.

(E) Example symmetric object in a symmetry detection task.

(F) Hypothesized neural computations for symmetry detection. Following multiple object normalization, the response to an object containing repeated parts is equal the response to the individual part, whereas the response to an object containing two different parts will lie along the line joining the objects with the two parts repeating. This causes symmetric objects to stand apart (red lines) and asymmetric objects to come closer due to mixing (*blue lines*). Thus, the visual homogeneity for symmetric objects (VH_s) will be larger than for asymmetric objects (VH_a). *Inset:* Schematic distance from center for symmetric objects (red) and asymmetric objects (blue).

(G) *Behavioral predictions for VH.* If visual homogeneity (VH) is a decision variable in visual search and symmetry detection tasks, then response times (RT) must be largest for displays with VH close to the decision boundary. This predicts opposite correlations between response time and VH for the present/absent or symmetry/asymmetry judgements. It also predicts zero overall correlation between VH and RT.

(H) *Neural predictions for VH. Left:* Correlation between brain activations and VH for two hypothetical brain regions. In the VH-encoding region, brain activations should be positively correlated with VH. In any region that encodes task difficulty as indexed by response time, brain activity should show no correlation since VH itself is uncorrelated with RT (see Panel G). *Right:* Correlation between brain activations and RT. Since VH is uncorrelated with RT overall, the region VH should show little or no correlation, whereas the regions encoding task difficulty would show a positive correlation.

RESULTS

In Experiments 1-2, we investigated whether visual homogeneity computations could explain decisions about targets being present or absent in an array. Since visual homogeneity requires measuring distance in perceptual space, we set out to first characterize the underlying representation of a set of natural objects using measurements of perceptual dissimilarity.

Measuring perceptual space for natural objects

In Experiment 1, 16 human participants viewed arrays made from a set of 32 grayscale natural objects, with an oddball on the left or right (Figure 2A), and had to indicate the side on which the oddball appeared using a key press. Participants were highly accurate and consistent in their responses during this task (accuracy, mean ± sd: 98.8 ± 0.9%; correlation between mean response times of even- and odd-numbered participants: $r = 0.91$, $p < 0.0001$ across all ${}^{32}C_2 = 496$ object pairs). The reciprocal of response time is a measure of perceptual distance (or dissimilarity) between the two images (Arun, 2012). To visualize the underlying object representation, we performed a multidimensional scaling analysis, which embeds objects in a multidimensional space such that their pairwise dissimilarities match the experimentally observed dissimilarities (see Methods). The resulting two-dimensional embedding of all objects is shown in Figure 2B. In the resulting plot, nearby objects correspond to hard searches, and far away objects correspond to easy searches. Such representations reconstructed from behavioural data closely match population neural responses in high-level visual areas (Op de Beeck et al., 2001; Sripati and Olson, 2010). To capture the object

representation accurately, we took the multidimensional embedding of all objects and treated the values along each dimension as the responses of an individual artificial neuron. We selected the number of dimensions in the multidimensional embedding so that the correlation between the observed and embedding dissimilarities matches the noise ceiling in the data. Subsequently, we averaged these single object responses to obtain responses to larger visual search arrays, as detailed below.

Visual homogeneity predicts target present/absent judgments (Experiments 1-2)

Having characterized the underlying perceptual representation for single objects, we set out to investigate whether target present/absent responses during visual search can be explained using this representation. In Experiment 2, 16 human participants viewed an array of items on each trial, and indicated using a key press whether there was an oddball target present or not (Figure 2C). This task was performed inside an MRI scanner to simultaneously observe both brain activity and behaviour. Participants were highly accurate and consistent in their responses (accuracy, mean ± sd: 95 ± 3%; correlation between average response times of even- and odd-numbered participants: $r = 0.86$, $p < 0.0001$ across 32 target-present searches, $r = 0.63$, $p < 0.001$ across 32 target-absent searches).

Next we set out to predict the responses to target-present and target-absent search displays containing these objects. We first took the object coordinates returned by multidimensional scaling in Experiment 1 as neural responses of multiple neurons. We then used a well-known principle of object representations in high-level visual areas: the response to multiple objects is the average of the single object responses (Zoccolan

et al., 2005; Agrawal et al., 2020). Thus, we took the response vector for a target-present array to be the average of the response vectors of the target and distractor (Figure 2D). Likewise, we took the response vector for a target-absent array to be equal to the response vector of the single item. We then asked if there is any point in this multidimensional representation such that distances from this point to the target-present and target-absent response vectors can accurately predict the target-present and target-absent response times with a positive and negative correlation respectively (see Methods). Note that this model has as many free parameters as the coordinates of this unknown point or center in multidimensional space, and this model can simultaneously predict both target-present and target-absent judgments. We used nonlinear optimization to find the coordinates of the center to best match the data (see Methods).

We denoted the distance of each display to the optimized center as the visual homogeneity. As expected, the visual homogeneity of target-present arrays was significantly smaller than target-absent arrays (Figure 2E). The resulting model predictions are shown in Figure 2F-G. The response times for target-present searches were positively correlated with visual homogeneity ($r = 0.68$, $p < 0.0001$; Figure 2F). By contrast, the response times for target-absent searches were negatively correlated with visual homogeneity ($r = -0.71$, $p < 0.0001$; Figure 2G). This is exactly as predicted if visual homogeneity is the underlying decision variable (Figure 1G). We note that the range of visual homogeneity values for target-present and target-absent searches do overlap, suggesting that visual homogeneity contributes but does not fully determine task performance. Rather, we suggest that visual homogeneity provides a useful and

initial first guess at the presence or absence of a target, which can be refined further through detailed scrutiny.

To confirm that the above model fits are not due to overfitting, we performed a leave-one-out cross validation analysis, where we left out all target-present and target-absent searches involving a particular image, and then predicted these searches by calculating visual homogeneity. This too yielded similar positive and negative correlations ($r = 0.63$, $p < 0.0001$ for target-present, $r = -0.63$, $p < 0.001$ for target-absent).

These findings are non-trivial for several reasons. First, we have shown that a specific decision variable, computed over an underlying object representation, can be used to make target present/absent judgements, without necessarily knowing the precise features of the target or distractor. Second, we have identified a specific image property that explains why target-absent response times vary so systematically. If target-distractor dissimilarity were the sole determining factor in visual search, it would predict no systematic variation in target-absent searches since the target-distractor dissimilarity is zero. Our results elucidate this puzzle by showing that visual homogeneity varies systematically across images, and that this explains systematic variation in target-absent response times. To the best of our knowledge, our model provides for the first time a unified explanation to explain how target present/absent judgements might be made in a generic visual search task.

We performed several additional analyses to validate these results and confirm their generality. First, if heterogeneous displays elicit similar neural responses due to mixing, then their average distance to other objects must be related to their visual

homogeneity. We confirmed that this was indeed the case, suggesting that the average distance of an object from all other objects in visual search can predict visual homogeneity (Section S1). Second, the above analysis was based on taking the neural response to oddball arrays to be the average of the target and distractor responses. To confirm that averaging was indeed optimal, we repeated the above analysis by assuming a range of relative weights between the target and distractor. The best correlation was obtained for almost equal weights in the lateral occipital (LO) region, consistent with averaging and its role in the underlying perceptual representation (Section S1). Third, we performed several additional experiments on a larger set of natural objects as well as on silhouette shapes. In all cases, present/absent responses were explained using visual homogeneity (Section S2).

In sum, we conclude that visual homogeneity can explain oddball target present/absent judgements during visual search.

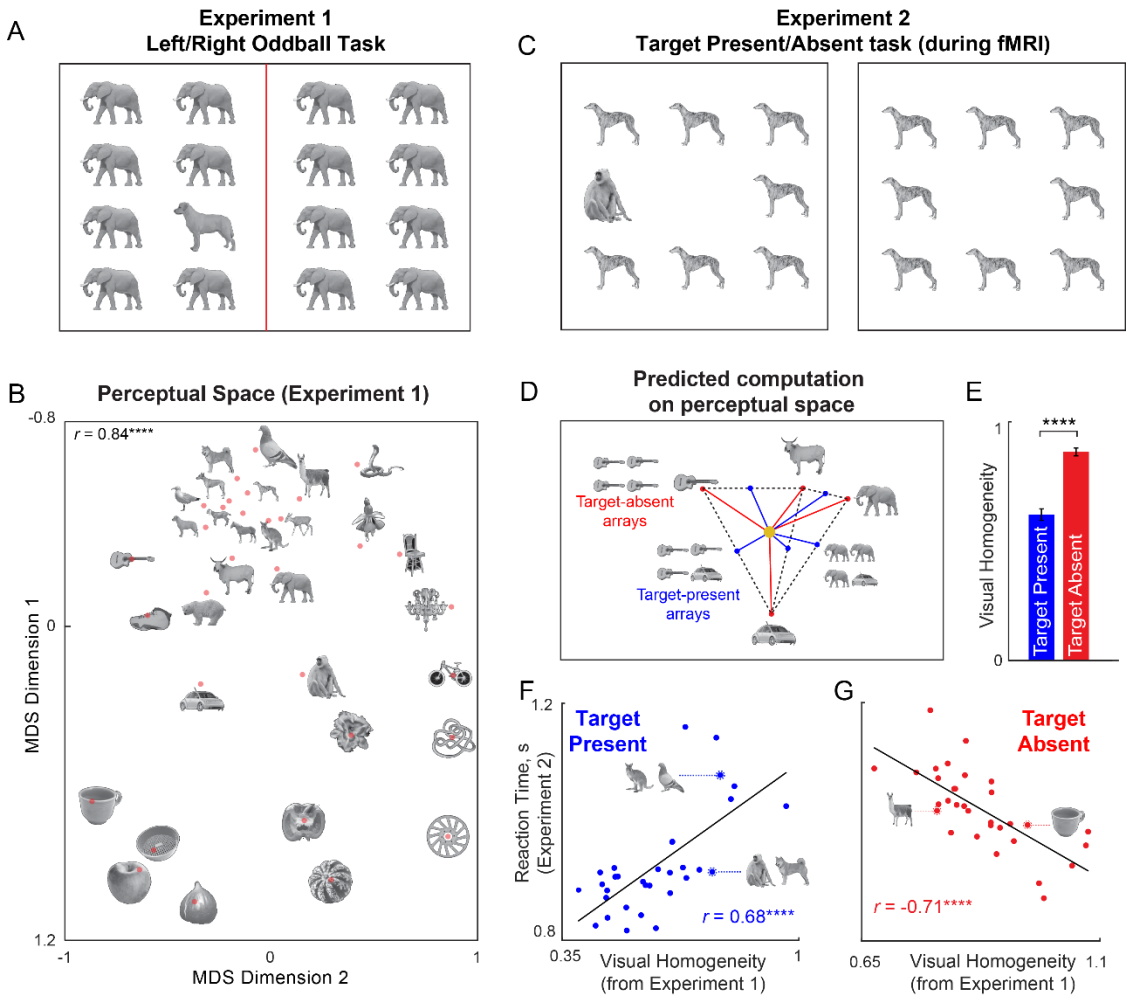

Figure 2. Visual homogeneity predicts target present/absent responses

(A) Example search array in an oddball search task (Experiment 1). Participants viewed an array containing identical items except for an oddball present either on the left or right side, and had to indicate using a key press which side the oddball appeared. The reciprocal of average search time was taken as the perceptual distance between the target and distractor items. We measured all possible pairwise distances for 32 grayscale natural objects in this manner.

(B) Perceptual space reconstructed using multidimensional scaling performed on the pairwise perceptual dissimilarities. In the resulting plot, nearby objects represent hard searches, and far away objects represent easy searches. Some images are shown at a small size due to space constraints; in the actual experiment, all objects were equated to have the same longer dimension. The correlation on the top right indicates the match between the distances in the 2D plot with the observed pairwise distances (**** is $p < 0.00005$).

(C) Example display from Experiment 2. Participants performed this task inside the scanner. On each trial, they had to indicate whether an oddball target is present or absent using a key press.

(D) Predicted response to target-present and target-absent arrays, using the principle that the neural response to multiple items is the average of the individual item responses. This predicts that target-present arrays become similar

due to mixing of responses, whereas target-absent arrays stand apart. Consequently, these two types of displays can be distinguished using their distance to a central point in this space. We define this distance as visual homogeneity, and it is obtained by finding the optimum center that maximizes the difference in correlations with response times (see Methods).

(E) Mean visual homogeneity relative to the optimum center for target-present and target-absent displays. Error bars represent s.e.m across all displays. Asterisks represent statistical significance (**** is $p < 0.00005$, unpaired rank-sum test comparing visual homogeneity for 32 target-absent and 32 target-present arrays).

(F) Response time for target-present searches in Experiment 2 plotted against visual homogeneity calculated from Experiment 1. Asterisks represent statistical significance of the correlation (**** is $p < 0.00005$). Note that a single model is fit to find the optimum center in representational space that predicts the response times for both target-present and target-absent searches.

(G) Response time for target-absent searches in Experiment 2 plotted against visual homogeneity calculated from Experiment 1. Asterisks represent statistical significance of the correlation (**** is $p < 0.00005$).

Visual homogeneity predicts same/different responses

We have proposed that visual homogeneity can be used to solve any task that requires discriminating between homogeneous and heterogeneous displays. In Experiments 1-2, we have shown that visual homogeneity predicts target present/absent responses in visual search. We performed an additional experiment to assess whether visual homogeneity can be used to solve an entirely different task, namely a same-different task. In this task, participants have to indicate whether two items are the same or different. We note that instructions to participants for the same/different task ("you have to indicate if the two items are same or different") are quite different from the visual search task ("you have to indicate whether an oddball target is present or absent"). Yet both tasks involve discriminating between homogeneous and heterogeneous displays. We therefore predicted that "same" responses would be correlated with target-absent judgements and "different" responses would be correlated with target-present judgements. Remarkably, this was indeed the case (Section S3), demonstrating that same/different responses can also be predicted using visual homogeneity.

Visual homogeneity is independent of experimental context

In the above analyses, visual homogeneity was calculated for each display as its distance from an optimum center in perceptual space. This raises the possibility that visual homogeneity could be modified depending on experimental context since it could depend on the set of objects relative to which the visual homogeneity is computed. We performed a number of experiments to evaluate this possibility: we found that target-

absent response times, which index visual homogeneity, are unaffected by a variety of experimental context manipulations (Section S4).

We therefore propose that visual homogeneity is an image-computable property that remains stable across tasks. Moreover, it can be empirically estimated as the reciprocal of the target-absent response times in a visual search task.

A localized brain region encodes visual homogeneity (Experiment 2)

So far, we have found that target present/absent response times had opposite correlations with visual homogeneity (Figure 2F-G), suggesting that visual homogeneity is a possible decision variable for this task. Therefore, we reasoned that visual homogeneity may be localized to specific brain regions, such as in the visual or prefrontal cortices. Since the task in Experiment 2 was performed by participants inside an MRI scanner, we set out to investigate this issue by analyzing their brain activations.

We estimated brain activations in each voxel for individual target-present and target-absent search arrays (see Methods). To identify the brain regions whose activations correlated with visual homogeneity, we performed a whole-brain searchlight analysis. For each voxel, we calculated the mean activity in a 3x3x3 volume centered on that voxel (averaged across voxels and participants) for each present/absent search display, and calculated its correlation with visual homogeneity predictions derived from behavior (see Methods). The resulting map is shown in Figure 3A. Visual homogeneity was encoded in a highly localized region just anterior of the lateral occipital (LO) region, with additional weak activations in the parietal and frontal regions. To compare these trends across key visual regions, we calculated the correlation between mean activation

and visual homogeneity for each region. This revealed visual homogeneity to be encoded strongly in this region VH, and only weakly in other visual regions (Figure 3D).

To ensure that the high match between visual homogeneity and neural activations in the VH region is not due to an artefact of voxel selection, we performed subject-level analysis (Section S5). We repeated the searchlight analysis for each subject and defined VH region for each subject. We find this VH region consistently anterior to the LO region in each subject. Next, we divided participants into two groups, and repeated the brain-wide searchlight analysis. Importantly, the match between mean activation and visual homogeneity remained significant even when the VH region was defined using one group of participants and the correlation was calculated using the mean activations of the other group (Section S5).

To confirm that neural activations in VH region are not driven by other cognitive processes linked to response time, such as attention, we performed a whole-brain searchlight analysis using response times across both target-present and target-absent searches. Proceeding as before, we calculated the correlation between mean activations to the target-present, target-absent and all displays with the respective response times. The resulting maps show that mean activations in the VH region are uncorrelated with response times overall (Section S5). By contrast, activations in EVC and LO are negatively correlated with response times, suggesting that faster responses are driven by higher activation of these areas. Finally, mean activation of parietal and prefrontal regions were strongly correlated with response times, consistent with their role in attentional modulation (Section S5).

Object representations in LO match with visual search dissimilarities

To investigate the neural space on which visual homogeneity is being computed, we performed a dissimilarity analysis. Since target-absent displays contain multiple instances of a single item, we took the neural response to target-absent displays as a proxy for the response to single items. For each pair of objects, we took the neural activations in a 3x3x3 neighborhood centered around a given voxel and calculated the Euclidean distance between the two 27-dimensional response vectors (averaged across participants). In this manner, we calculated the neural dissimilarity for all ${}^{32}C_2$ = 496 pairs of objects used in the experiment, and calculated the correlation between the neural dissimilarity in each local neighborhood and the perceptual dissimilarities for the same objects measured using oddball search in Experiment 1. The resulting map is shown in Figure 3B. It can be seen that perceptual dissimilarities from visual search are best correlated in the lateral occipital region, consistent with previous studies (Figure 3E). To compare these trends across key visual regions, we performed this analysis for early visual cortex (EVC), area V4, LO and for the newly identified region VH (average MNI coordinates (x, y, z): (-48, -59, -6) with 111 voxels in the left hemisphere; (49, -56, -7) with 60 voxels in the right hemisphere). Perceptual dissimilarities matched best with neural dissimilarities in LO compared to the other visual regions (Figure 3E). We conclude that neural representations in LO match with perceptual space. This is concordant with many previous studies (Haushofer et al., 2008; Kriegeskorte et al., 2008; Agrawal et al., 2020; Storrs et al., 2021; Ayzenberg et al., 2022).

Equal weights for target and distractor in target-present array responses

In the preceding sections, visual homogeneity was calculated using behavioural experiments assuming a neural representation that gives equal weights to the target and distractor. We tested this assumption experimentally by asking whether neural responses to target-present displays can be predicted using the response to the target and distractor items separately. The resulting maps revealed that target-present arrays were accurately predicted as a linear sum of the constituent items, with roughly equal weights for the target and distractor (Section S5).

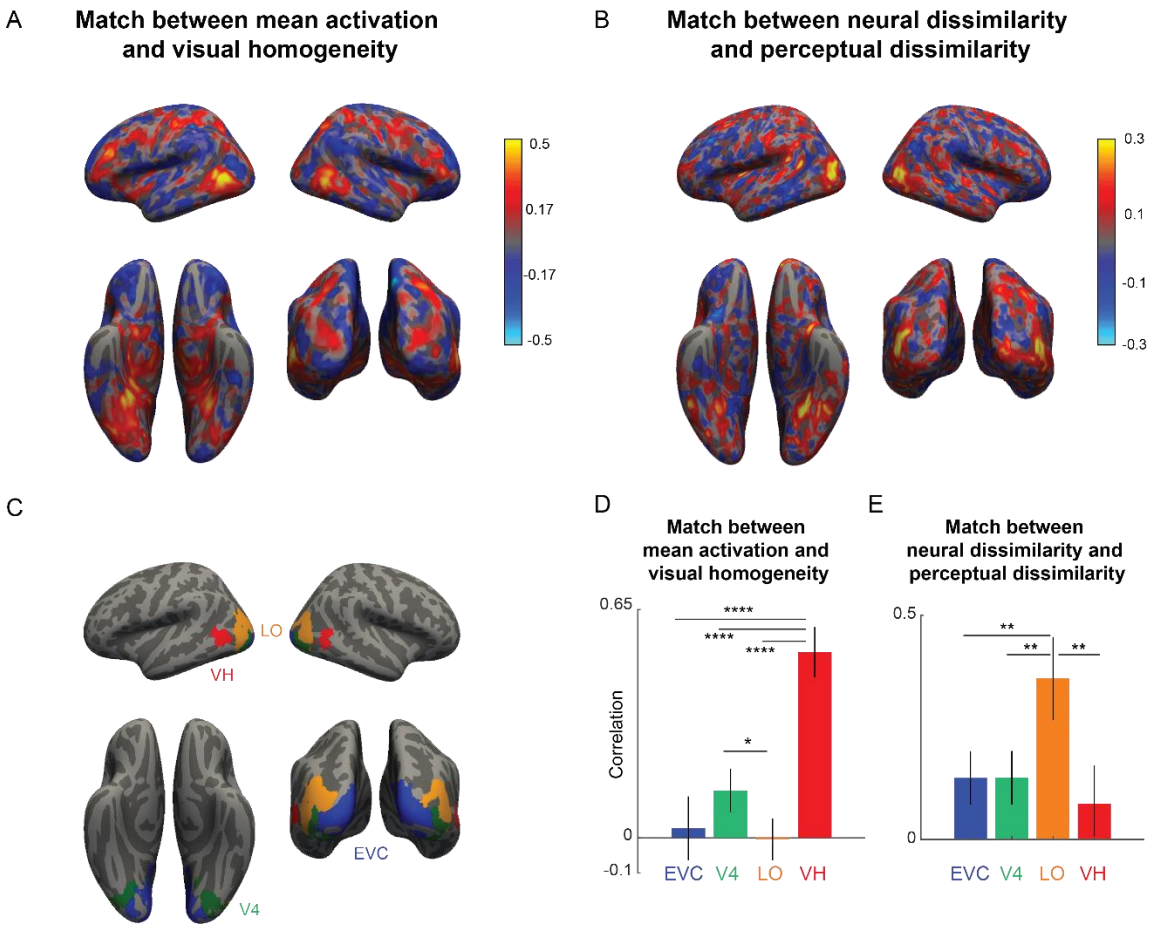

Figure 3: A localized brain region encodes visual homogeneity

A. Searchlight map showing the correlation between mean activation in each 3x3x3 voxel neighborhood and visual homogeneity.
B. Searchlight map showing the correlation between neural dissimilarity in each 3x3x3 voxel neighborhood and perceptual dissimilarity measured in visual search.
C. Key visual regions identified using standard anatomical masks: early visual cortex (EVC), area V4, lateral occipital (LO) region. The visual homogeneity (VH) region was identified using the searchlight map in Panel A.
D. Correlation between the mean activation and visual homogeneity in key visual regions EVC, V4, LO and VH. Error bars represent standard deviation of the correlation obtained using a boostrap process, by repeatedly sampling participants with replacement for 10,000 times. Asterisks represent statistical significance, estimated by calculating the fraction of bootstrap samples in which the observed trend was violated (* is $p < 0.05$, ** is $p < 0.01$, **** is $p < 0.0001$).
E. Correlation between neural dissimilarity in key visual regions with perceptual dissimilarity. Error bars represent the standard deviation of correlation obtained using a bootstrap process, by repeatedly sampling participants with replacement 10,000 times. Asterisks represent statistical significance, estimated by calculating the fraction of bootstrap samples in which the observed trend was violated (** is $p < 0.001$).

Visual homogeneity predicts symmetry perception (Experiments 3-4)

The preceding sections show that visual homogeneity predicts target present/absent responses as well same/different responses. We have proposed that visual homogeneity can be used to solve any task that involves discriminating homogeneous and heterogeneous displays. In Experiments 3 & 4, we extend the generality of these findings to an entirely different task, namely symmetry perception. Here, asymmetric objects are akin to heterogeneous displays whereas symmetric objects are like homogeneous displays.

In Experiment 3, we measured perceptual dissimilarities for a set of 64 objects (32 symmetric, 32 asymmetric objects) made from a common set of parts. On each trial, participants viewed a search array with identical items except for one oddball, and had to indicate the side (left/right) on which the oddball appeared using a key press. An example search array is shown in Figure 4A. Participants performed searches involving all possible ${}^{64}C_2$ = 2,016 pairs of objects. Participants made highly accurate and consistent responses on this task (accuracy, mean ± sd: 98.5 ± 1.33%; correlation between average response times from even- and odd-numbered subjects: $r = 0.88$, $p < 0.0001$ across 2,016 searches). As before, we took the perceptual dissimilarity between each pair of objects to be the reciprocal of the average response time for displays with either item as target and the other as distractor. To visualize the underlying object representation, we performed a multidimensional scaling analysis, which embeds objects in a multidimensional space such that their pairwise dissimilarities match the experimentally observed dissimilarities. The resulting plot for two dimensions is shown in Figure 4B, where nearby objects correspond to similar searches. It can be seen that

symmetric objects are generally more spread apart than asymmetric objects, suggesting that visual homogeneity could discriminate between symmetric and asymmetric objects.

In Experiment 4, we tested this prediction experimentally using a symmetry detection task that was performed by participants inside an MRI scanner. On each trial, participants viewed a briefly presented object, and had to indicate whether the object was symmetric or asymmetric using a key press (Figure 4C). Participants made accurate and consistent responses in this task (accuracy, mean ± sd: 97.7 ± 1.7%; correlation between response times of odd- and even-numbered participants: $r = 0.47$, $p < 0.0001$).

To investigate whether visual homogeneity can be used to predict symmetry judgements, we took the embedding of all objects from Experiment 3, and asked whether there was a center in this multidimensional space such that the distance of each object to this center would be oppositely correlated with response times for symmetric and asymmetric objects (see Methods). The resulting model predictions are shown in Figure 4E-G. As predicted, visual homogeneity was significantly larger for symmetric compared to asymmetric objects (visual homogeneity, mean ± sd: 0.60 ± 0.24 s^{-1} for asymmetric objects; 0.76 ± 0.29 s^{-1} for symmetric objects; $p < 0.05$, rank-sum test; Figure 4E). For asymmetric objects, symmetry detection response times were positively correlated with visual homogeneity ($r = 0.56$, $p < 0.001$; Figure 4F). By contrast, for symmetric objects, response times were negatively correlated with visual homogeneity ($r = -0.39$, $p < 0.05$; Figure 4G). These patterns are exactly as expected if visual homogeneity was the underlying decision variable for symmetry detection. However, we note that the range of visual homogeneity values for asymmetric and

symmetric objects do overlap, suggesting that visual homogeneity contributes but does not fully determine task performance. We therefore propose that visual homogeneity provides a useful and initial first guess at symmetry in an image, which can be refined further through detailed scrutiny.

To confirm that the above model fits are not due to overfitting, we performed a leave-one-out cross validation analysis, where we left out one object at a time, and then calculated its visual homogeneity. This too yielded similar correlations ($r = 0.44$ for asymmetric, $r = -0.39$ for symmetric objects, $p < 0.05$ in both cases).

In sum, we conclude that visual homogeneity can predict symmetry perception. Taken together, these experiments demonstrate that the same computation (distance from a center) explains two disparate generic visual tasks: symmetry perception and visual search.

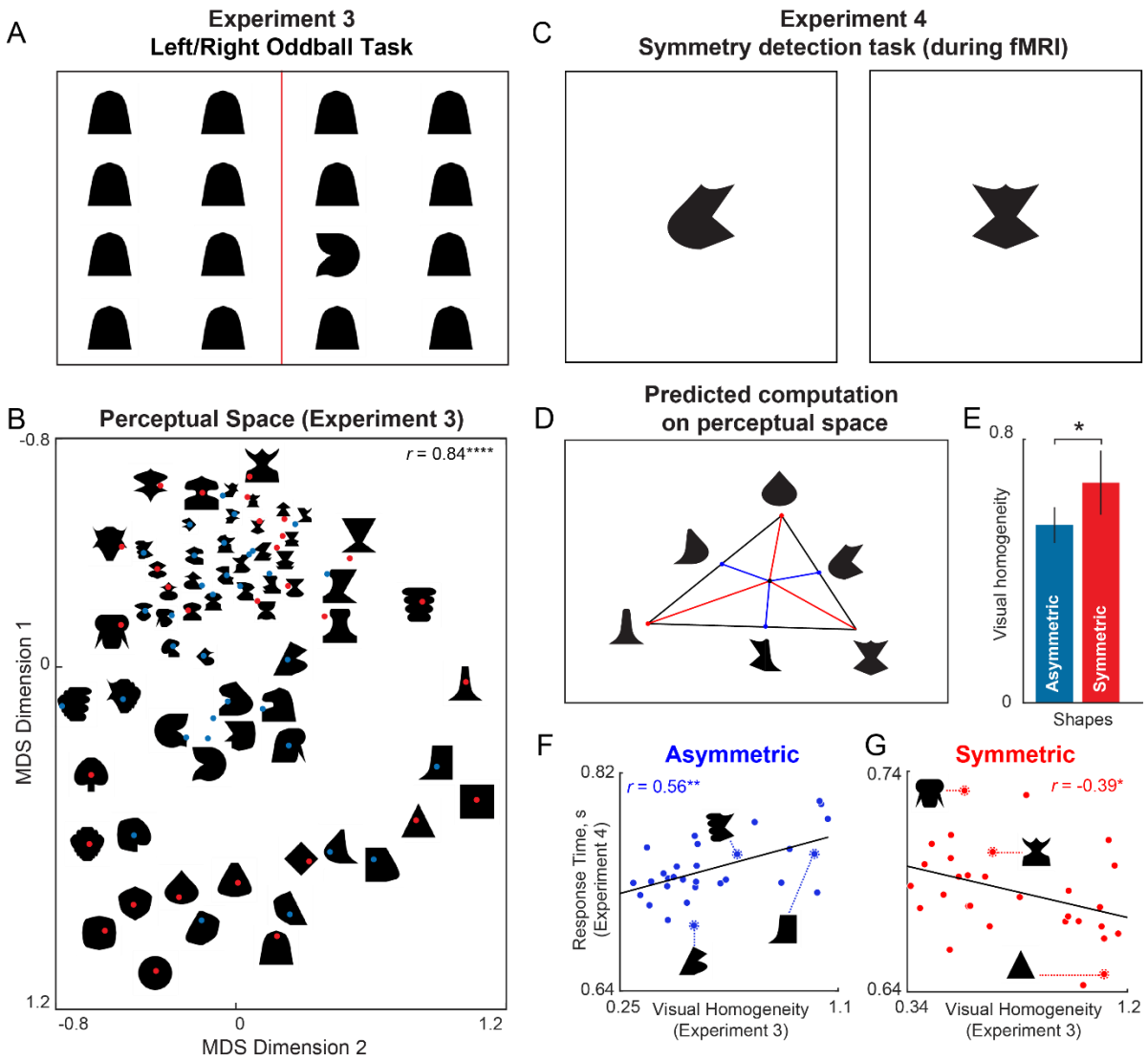

Figure 4. Visual homogeneity predicts symmetry perception

(A) Example search array in Experiment 3. Participants viewed an array containing identical items except for an oddball present either on the left or right side, and had to indicate using a key press which side the oddball appeared. The reciprocal of average search time was taken as the perceptual distance between the target and distractor items. We measured all possible pairwise distances for 64 objects (32 symmetric, 32 asymmetric) in this manner.

(B) Perceptual space reconstructed using multidimensional scaling performed on the pairwise perceptual dissimilarities. In the resulting plot, nearby objects represent hard searches, and far away objects represent easy searches. Some images are shown at a small size due to space constraints; in the actual experiment, all objects were equated to have the same longer dimension. The correlation on the *top right* indicates the match between the distances in the 2D plot with the observed pairwise distances (**** is $p < 0.00005$).

(C) Two example displays from Experiment 4. Participants had to indicate whether the object is symmetric or asymmetric using a key press.

(D) Using the perceptual representation of symmetric and asymmetric objects from Experiment 3, we reasoned that they can be distinguished using their distance to a center in perceptual space. The coordinates of this center are optimized to maximize the match to the observed symmetry detection times.

(E) Visual homogeneity relative to the optimum center for asymmetric and symmetric objects. Error bar represents s.e.m. across images. Asterisks represent statistical significance (* is $p < 0.05$, unpaired rank-sum test comparing visual homogeneity for 32 symmetric and 32 asymmetric objects).

(F) Response time for asymmetric objects in Experiment 4 plotted against visual homogeneity calculated from Experiment 3. Asterisks represent statistical significance of the correlation (** is $p < 0.001$).

(G) Response time for symmetric objects in Experiment 4 plotted against visual homogeneity calculated from Experiment 3. Asterisks represent statistical significance of the correlation (* is $p < 0.05$).

Visual homogeneity is encoded by the VH region during symmetry detection

If visual homogeneity is a decision variable for symmetry detection, it could be localized to specific regions in the brain. To investigate this issue, we analyzed the brain activations of participants in Experiment 4.

To investigate the neural substrates of visual homogeneity, we performed a searchlight analysis. For each voxel, we calculated the correlation between mean activations in a 3x3x3 voxel neighborhood and visual homogeneity. This revealed a localized region in the visual cortex as well as some parietal regions where this correlation attained a maximum (Figure 5A). This VH region (average MNI coordinates (x, y, z): (-57, -56, -8) with 93 voxels in the left hemisphere; (58, -50, -8) with 73 voxels in the right hemisphere) overlaps with VH region defined during visual search in Experiment 3 (for a detailed comparison, see Section S7).

We note that it is not straightforward to interpret the overlap between the VH regions identified in Experiments 2 & 4, since images in this experiment (Experiment 4) were presented at the fovea, whereas the visual search arrays in Experiment 2 contained items exclusively in the periphery. These experiments were also performed on different sets of participants, so the differences in overlap could be due to individual variability. Indeed, considerable individual variability has been observed in the location of category-selective regions such as FFA (Weiner and Grill-Spector, 2012) and VWFA (Glezer and Riesenhuber, 2013). We propose that testing the same participants on both search and symmetry tasks would reveal VH regions with even larger overlap.

To confirm that neural activations in VH region are not driven by other cognitive processes linked to response time, such as attention, we performed a whole-brain

searchlight analysis using response times across both symmetric and asymmetric objects. This revealed that mean activations in the VH region were poorly correlated with response times overall, whereas other parietal and prefrontal regions strongly correlated with response times consistent with their role in attentional modulation (Section S6).

To investigate the perceptual representation that is being used for visual homogeneity computations, we performed a neural dissimilarity analysis. For each pair of objects, we took the neural activations in a 3x3x3 neighborhood centered around a given voxel and calculated the Euclidean distance between the two 27-dimensional response vectors. In this manner, we calculated the neural dissimilarity for all ${}^{64}C_2$ = 2,016 pairs of objects used in the experiment, and calculated the correlation between the neural dissimilarity in each local neighborhood and the perceptual dissimilarities for the same objects measured using oddball search in Experiment 3. The resulting map is shown in Figure 5B. The match between neural and perceptual dissimilarity was maximum in the lateral occipital region (Figure 5B).

To compare these trends for key visual regions, we repeated this analysis for anatomically defined regions of interest in the visual cortex: early visual cortex (EVC), area V4, the lateral occipital (LO) region, and the VH region defined based on the searchlight map in Figure 5A. These regions are depicted in Figure 5C. We then asked how mean activations in each of these regions is correlated with visual homogeneity. This revealed that the match with visual homogeneity is best in the VH region compared to the other regions (Figure 5D). To further confirm that visual homogeneity is encoded in a localized region in the symmetry task, we repeated the searchlight analysis on two

independent subgroups of participants. This revealed highly similar regions in both groups (Section S6).

Finally, we compared neural dissimilarities and perceptual dissimilarities in each region as before. This revealed that perceptual dissimilarities (measured from Experiment 3, during visual search) matched best with the LO region (Figure 5E), suggesting that object representations in LO are the basis for visual homogeneity computations in the VH region.

In sum, our results suggest that visual homogeneity is encoded by the VH region, using object representations present in the adjoining LO region.

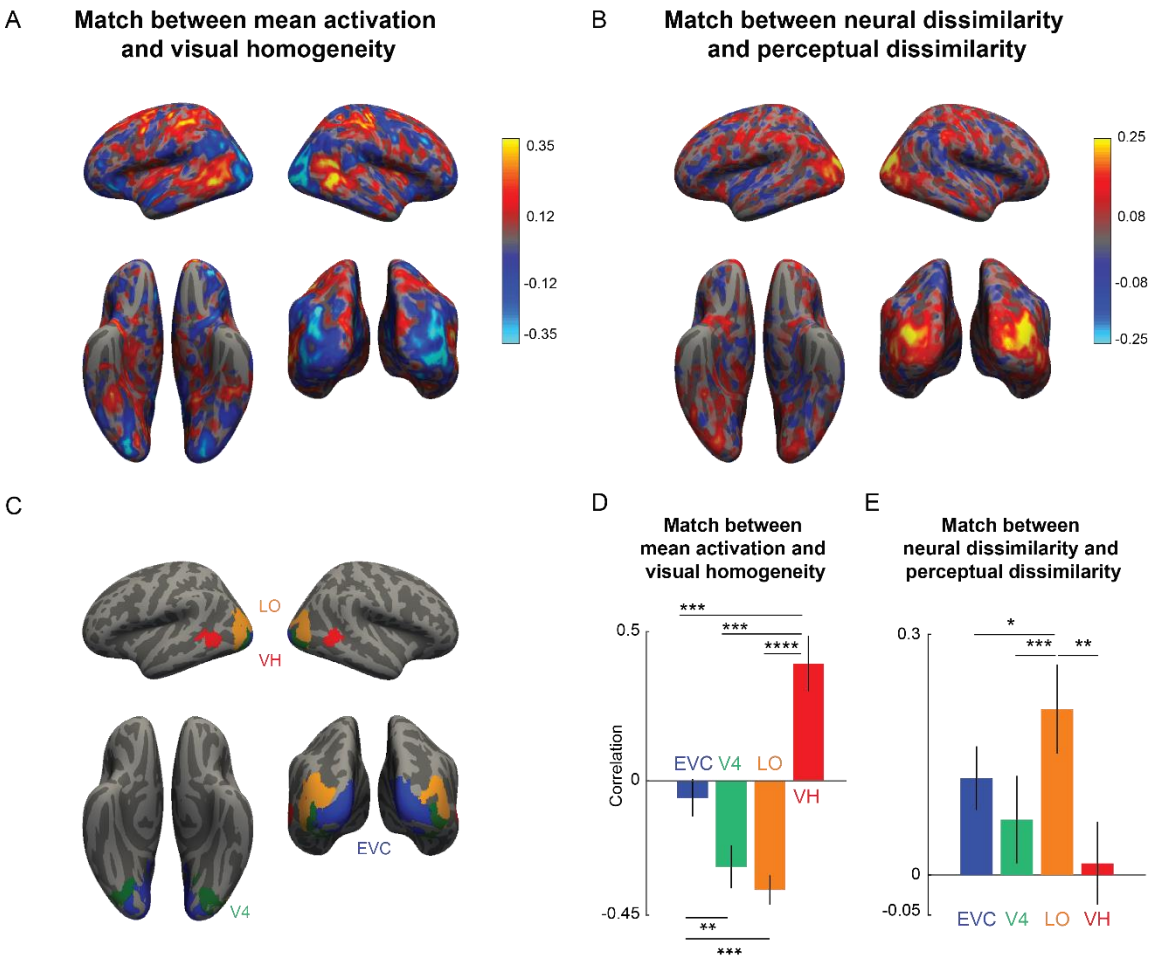

Figure 5: Brain region encoding visual homogeneity during symmetry detection

(A) Searchlight map showing the correlation between mean activation in each 3x3x3 voxel neighborhood and visual homogeneity.

(B) Searchlight map showing the correlation between neural dissimilarity in each 3x3x3 voxel neighborhood and perceptual dissimilarity measured in visual search.

(C) Key visual regions identified using standard anatomical masks: early visual cortex (EVC), area V4, Lateral occipital (LO) region. The visual homogeneity (VH) region was identified using searchlight map in Panel A.

(D) Correlation between the mean activation and visual homogeneity in key visual regions EVC, V4, LO and VH. Error bars represent standard deviation of the correlation obtained using a boostrap process, by repeatedly sampling participants with replacement for 10,000 times. Asterisks represent statistical significance, estimated by calculating the fraction of bootstrap samples in which the observed trend was violated (* is $p < 0.05$, ** is $p < 0.01$, **** is $p < 0.0001$).

(E) Correlation between neural dissimilarity in key visual regions with perceptual dissimilarity. Error bars represent the standard deviation of correlation obtained using a bootstrap process, by repeatedly sampling participants with replacement 10,000 times. Asterisks represent statistical significance, estimated by calculating the fraction of bootstrap samples in which the observed trend was violated (** is $p < 0.001$).

Target-absent responses predict symmetry detection

So far, we have shown that visual homogeneity predicts target present/absent responses in visual search as well as symmetry detection responses. These results suggest a direct empirical link between these two tasks. Specifically, since target-absent response time is inversely correlated with visual homogeneity, we can take its reciprocal as an estimate of visual homogeneity. This in turn predicts opposite correlations between symmetry detection times and reciprocal of target-absent response time: in other words, we should see a positive correlation for asymmetric objects, and a negative correlation for symmetric objects. We confirmed these predictions using additional experiments (Section S8). These results reconfirm that a common decision variable, visual homogeneity, drives both target present/absent and symmetry judgements.

Visual homogeneity explains animate categorization

Since visual homogeneity is always calculated relative to a location in perceptual space, we reasoned that shifting this center towards a particular object category would make it a decision variable for object categorization. To test this prediction, we reanalyzed data from a previous study in which participants had to categorize images as belonging to three hierarchical categories: animals, dogs or Labradors (Mohan and Arun, 2012). By adjusting the center of the perceptual space measured using visual search, we were able to predict categorization responses for all three categories (Section S9). We further reasoned that, if the optimum center for animal/dog/Labrador categorization is close to the default center in perceptual space that predicts target present/absent judgements, then even the default visual homogeneity, as indexed by

the reciprocal of target-absent search time, should predict categorization responses. Interestingly, this was indeed the case (Section S9). We conclude that, at least for the categories tested, visual homogeneity computations can serve as a decision variable for object categorization.

DISCUSSION

Here, we investigated three disparate visual tasks: detecting whether an oddball is present in a search array, deciding if two items are same or different, and judging whether an object is symmetric/asymmetric. Although these tasks are superficially quite different, they all involve discriminating between homogeneous and heterogeneous displays. We defined a new image property computable from the underlying perceptual representation, namely visual homogeneity, that can be used to solve these tasks. We found that visual homogeneity can predict response times in all three tasks. Finally we have found that visual homogeneity estimated from behavior is correlated with mean activations anterior to the lateral occipital region in both visual search and symmetry tasks. This is indicative but not conclusive evidence for a specific function to this part of the high-level visual cortex. Below we discuss these findings in relation to the existing literature.

Visual homogeneity unifies visual search, same-different and symmetry tasks

Our main finding, that a single decision variable (visual homogeneity) can be used to solve three disparate visual tasks (visual search, same/different and symmetry detection) is novel to the best of our knowledge. This finding is interesting and important because it establishes a close correspondence between all three tasks, and explains some unresolved puzzles in each of these tasks, as detailed below.

First, with regards to visual search, theoretical accounts of search are based on signal detection theory (Verghese, 2001; Wolfe and Horowitz, 2017), but define the signal only for specific target-distractor pairs. By contrast, the task of detecting whether

an oddball item is present requires a more general decision rule that has not been identified. Our results suggest that visual homogeneity is the underlying decision variable. In visual search, target-present search times are widely believed to be driven by target-distractor similarity (Duncan and Humphreys, 1989; Wolfe and Horowitz, 2004). But target-absent search times also vary systematically for reasons that have not been elucidated previously. The slope of target-absent search times as a function of set size are typically twice the slope of target present searches (Wolfe, 1998). However this observation is based on averaging across many target-present searches. Since there is only one unique item in a target-absent search array, there must be some image property that causes systematic variations. Our results resolve this puzzle by showing that this systematic variation is driven by visual homogeneity. Finally, our findings also help explain why we sometimes know a target is present without knowing its exact location – this is because the underlying decision variable, visual homogeneity, arises in high-level visual areas with relatively coarse spatial information. However, we note that visual homogeneity computations described in this study do not completely explain all the variance observed in oddball search times in our study – rather they offer an important quantitative model that could explain the initial phase of target selection. We speculate that this initial phase could be shared by all forms of visual search (e.g. searching among non-identical distractors, memory-guided search, conjunction search), and these would be interesting possibilities for future work.

Second, with regards to same-different tasks, most theoretical accounts use signal detection theory but usually with reference to specific stimulus pairs (Nickerson, 1969; Petrov, 2009). It has long been observed that "different" responses become faster

with increasing target-distractor dissimilarity but this trend logically predicts that "same" responses, which have zero difference, should be the slowest (Nickerson, 1967, 1969). But in fact, "same" responses are faster than "different" responses. This puzzle has been resolved by assuming a separate decision rule for "same" judgements, making the overall decision process more complex (Petrov, 2009; Goulet, 2020). Our findings resolve this puzzle by identifying a novel variable, visual homogeneity, which can be used to implement a simple decision rule for making same/different responses. Our findings also explain why some images elicit faster "same" responses than others: this is due to image-to-image differences in visual homogeneity.

Third, with regard to symmetry detection, most theoretical accounts assume that symmetry is explicitly detected using symmetry detectors along particular axes (Wagemans, 1997; Bertamini and Makin, 2014). By contrast, our findings indicate an indirect mechanism for symmetry detection that does not invoke any special symmetry computations. We show that visual homogeneity computations can easily discriminate between symmetric and asymmetric objects. This is because symmetric objects have high visual homogeneity since they have repeated parts, whereas asymmetric objects have low visual homogeneity since they have disparate parts (Pramod and Arun, 2018). In a recent study, symmetry detection was explained by the average distance of objects relative to other objects (Pramod and Arun, 2018). This finding is consistent with ours since visual homogeneity is correlated with the average distance to other objects (Section S1). However, there is an important distinction between these two quantities. Visual homogeneity is an intrinsic image property, whereas the average distance of an object to other objects depends on the set of other objects on which the average is

being computed. Indeed, we have confirmed through additional experiments that visual homogeneity is independent of experimental context (Section S4). We speculate that visual homogeneity can explain many other aspects of symmetry perception, such as the relative strength of symmetries.

Visual homogeneity in other visual tasks

Our finding that visual homogeneity explains generic visual tasks has several important implications for visual tasks in general. First, we note that visual homogeneity can be easily extended to explain other generic tasks such as delayed match-to-sample tasks or n-back tasks, by taking the response to the test stimulus as being averaged with the sample-related information in working memory. In such tasks, visual homogeneity will be larger for sequences with repeated compared to non-repeated stimuli, and can easily be used to solve the task. Testing these possibilities will require comparing systematic variations in response times in these tasks across images, and measurements of perceptual space for calculating visual homogeneity.

Second, we note that visual homogeneity can also be extended to explain object categorization, if one assumes that the center in perceptual space for calculating visual homogeneity can be temporarily shifted to the center of an object category. In such tasks, visual homogeneity relative to the category center will be small for objects belonging to a category and large for objects outside the category, and can be used as a decision variable to solve categorization tasks. This idea is consistent with prevalent accounts of object categorization (Stewart and Morin, 2007; Ashby and Maddox, 2011; Mohan and Arun, 2012). Indeed, categorization response times can be explained using

perceptual distances to category and non-category items (Mohan and Arun, 2012). By reanalyzing data from this study, we have found that, at least for the animate categories tested, visual homogeneity can explain categorization responses (Section S9). However, this remains to be tested in a more general fashion across multiple object categories.

Neural encoding of visual homogeneity

We have found that visual homogeneity is encoded in a specific region of the brain, which we denote as region VH, in both visual search and symmetry detection tasks (Figure 3D, 5D). This finding is consistent with observations of norm-based encoding in IT neurons (Leopold et al., 2006) and in face recognition (Valentine, 1991; Rhodes and Jeffery, 2006; Carlin and Kriegeskorte, 2017). However, our findings are significant because they reveal a dedicated region in high-level visual cortex for solving generic visual tasks.

We have found that the VH region is located just anterior to the lateral occipital (LO) region, where neural dissimilarities match closely with perceptual dissimilarities (Figure 3E, 5E). Based on this proximity, we speculate that visual homogeneity computations are based on object representations in LO. However, confirming this prediction will require fine-grained recordings of neural activity in VH and LO. An interesting possibility for future studies would be to causally perturb brain activity separately in VH or LO using magnetic or electrical stimulation, if at all possible. A simple prediction would be that perturbing LO would distort the underlying representation, whereas perturbing VH would distort the underlying decision process.

We caution however that the results might not be so easily interpretable if visual homogeneity computations in VH are based on object representations in LO.

Recent observations from neural recordings in monkeys suggest that perceptual dissimilarities and visual homogeneity need not be encoded in separate regions. For instance, the overall magnitude of the population neural response of monkey inferior temporal (IT) cortex neurons was found to correlate with memorability (Jaegle et al., 2019). These results are consistent with encoding of visual homogeneity in these regions. However, we note that neural responses in IT cortex also predict perceptual dissimilarities (Op de Beeck et al., 2001; Sripati and Olson, 2010; Zhivago and Arun, 2014; Agrawal et al., 2020). Taken together, these findings suggest that visual homogeneity computations and the underlying perceptual representation could be interleaved within a single neural population, unlike in humans where we found separate regions. Indeed, in our study, the mean activations of the LO region were also correlated with visual homogeneity for symmetry detection (Figure 5A), but not for target present/absent search (Figure 3A). We speculate that perhaps visual homogeneity might be intermingled into the object representation in monkeys but separated into a dedicated region in humans. These are interesting possibilities for future work.

Although many previous studies have reported brain activations in the vicinity of the VH region, we are unaware of any study that has ascribed a specific function to this region. The localized activations in our study match closely with the location of the recently reported ventral stream attention module in both humans and monkeys (Sani et al., 2021). Previous studies have observed important differences in brain activations in this region, which can be explained using visual homogeneity, as detailed below.

First, previous studies have observed differences in brain activations for animate compared to inanimate objects in high-level visual areas which have typically included the newly defined VH region reported here (Bracci and Op de Beeck, 2015; Proklova et al., 2016; Thorat et al., 2019). These observations would be consistent with our findings if visual homogeneity is smaller for animate compared to inanimate objects, which would result in weaker activations for animate objects in region VH. Indeed, visual homogeneity, as indexed by the reciprocal of target-absent search time, is smaller for animate objects compared to inanimate objects (Section S9). Likewise, brain activations were weaker for animate objects compared to inanimate objects in region VH (average VH activations, mean ± sd across participants: 0.50 ± 0.61 for animate target-absent displays, 0.64 ± 0.59 for inanimate target-absent displays, $p < 0.05$, sign-rank test across participants). Based on this we speculate that visual homogeneity may partially explain the animacy organization of human ventral temporal cortex. Establishing this will require controlling animate/inanimate stimuli not only for shape but also for visual homogeneity.

Second, previous studies have reported larger brain activations for symmetric objects compared to asymmetric objects in the vicinity of this region (Sasaki et al., 2005; Van Meel et al., 2019). This can be explained by our finding that symmetric objects have larger visual homogeneity (Figure 4E), leading to activation of the VH region (Figure 5A). But the increased activations in previous studies were located in the V4 & LO regions, whereas we have found greater activations more anteriorly in the VH region. This difference could be due to the stimulus-related differences: both previous studies used dot patterns, which could appear more object-like when symmetric, leading

to more widespread differences in brain activation due to other visual processes like figure-ground segmentation (Van Meel et al., 2019). By contrast, both symmetric and asymmetric objects in our study are equally object-like. Resolving these discrepancies will require measuring visual homogeneity as well as behavioural performance during symmetry detection for dot patterns.

Finally, our results are consistent with the greater activity observed for objects with shared features observed in ventral temporal cortex during naming tasks (Taylor et al., 2012; Tyler et al., 2013). Our study extends these observations by demonstrating an empirical measure for shared feature (target-absent times in visual search) and encoding of this empirical measure into a localized region in object selective cortex across many tasks. We speculate that visual homogeneity may at least partially explain semantic concepts such as those described in these studies.

Relation to image memorability

We have defined a novel image property, visual homogeneity, which refers to the distance of a visual image to a central point in the underlying perceptual representation. It can be reliably estimated for each image as the inverse of the target-absent response time in a visual search task (Figure 2) and is independent of the immediate experimental context (Section S4).

The above observations raise the question of whether and how visual homogeneity might be related to image memorability. It has long been noted that faces that are rated as being distinctive or unusual are also easier to remember (Murdock, 1960; Valentine and Bruce, 1986a, 1986b; Valentine, 1991). Recent studies have

elucidated this observation by showing that there are specific image properties that predict image memorability (Bainbridge et al., 2017; Lukavský and Děchtěrenko, 2017; Rust and Mehrpour, 2020). However, image memorability, as elegantly summarized in a recent review (Rust and Mehrpour, 2020), could be driven by a number of both intrinsic and extrinsic factors. It would therefore be interesting to characterize how well visual homogeneity, empirically measured using target-absent visual search, can predict image memorability.

Conclusions

Taken together, our results show that a variety of generic visual tasks that require detecting feature-independent image properties, can be solved using visual homogeneity as a decision variable, which is localized to a specific region anterior to the lateral occipital region. While this does not explain all possible variations of these tasks, our study represents an important first step in terms of demonstrating a quantitative, falsifiable model and a localized neural substrate. These can serve as a basis to elucidate whether/how visual homogeneity computations might contribute to a variety of visual tasks.

MATERIALS AND METHODS

All participants had a normal or corrected-to-normal vision and gave informed consent to an experimental protocol approved by the Institutional Human Ethics Committee of the Indian Institute of Science (IHEC # 6-15092017). Participants provided written informed consent before each experiment and were monetarily compensated.

Experiment 1. Oddball detection for perceptual space (natural objects)

Participants. A total of 16 participants (8 males, 22 ± 2.8 years) participated in this experiment.

Stimuli. The stimulus set comprised a set of 32 grayscale natural objects (16 animate, 16 inanimate) presented against a black background.

Procedure. Participants performed an oddball detection task with a block of practice trials involving unrelated stimuli followed by the main task. Each trial began with a red fixation cross (diameter 0.5°) for 500 ms, followed by a 4 x 4 search array measuring 30° x 30° for 5 seconds or until a response was made. The search array always contained one oddball target and 15 identical distractors, with the target appearing equally often on the left or right. A vertical red line divided the screen equally into two halves to facilitate responses. Participants were asked to respond as quickly and as accurately as possible using a key press to indicate the side of the screen containing

the target ('Z' for left, M' for right). Incorrect trials were repeated later after a random number of other trials. Each participant completed 992 correct trials ($^{32}C_2$ object pairs x 2 repetitions with either image as target). The experiment was created using PsychoPy (Peirce et al., 2019) and ported to the online platform Pavlovia for collecting data.

Since stimulus size could vary with the monitor used by the online participants, we equated the stimulus size across participants using the ScreenScale function (https://doi.org/10.17605/OSF.IO/8FHQK). Each participant adjusted the size of a rectangle on the screen such that its size matched the physical dimensions of a credit card. All the stimuli presented were then scaled with the estimated scaling function to obtain the desired size in degrees of visual angle, assuming an average distance to screen of 60 cm.

Data Analysis. Response times faster than 0.3 seconds or slower than 3 seconds were removed from the data. This step removed only 1.25% of the data and improved the overall response time consistency, but did not qualitatively alter the results.

Characterizing perceptual space using multidimensional scaling. To characterize the perceptual space on which present/absent decisions are made, we took the inverse of the average response times (across trials and participants) for each image pair. This inverse of response time (i.e. 1/RT) represents the dissimilarity between the target and distractor (Arun, 2012), indexes the underlying salience signal in visual search (Sunder and Arun, 2016) and combines linearly across a variety of factors (Pramod and Arun, 2014, 2016; Jacob and Arun, 2020). Since there were 32 natural objects in the

experiment and all possible ($^{32}C_2$ = 496) pairwise searches in the experiment, we obtained 496 pairwise dissimilarities overall. To calculate target-present and target-absent array responses, we embedded these objects into a multidimensional space using multidimensional scaling analysis (*mdscale* function; MATLAB 2019). This analysis finds the n-dimensional coordinates for each object such that pairwise distances between objects best matches with the experimentally observed pairwise distances. We then treated the activations of objects along each dimension as the responses of a single artificial neuron, so that the response to target-present arrays could be computed as the average of the target and distractor responses.

Experiment 2. Target present-absent search during fMRI

Participants. A total of 16 subjects (11 males; age, mean ± sd: 25 ± 2.9 years) participated in this experiment. Participants with history of neurological or psychiatric disorders, or with metal implants or claustrophobia were excluded through screening questionnaires.

Procedure. Inside the scanner, participants performed a single run of a one-back task for functional localizers (block design, object vs scrambled objects), eight runs of the present-absent search task (event-related design), and an anatomical scan. The experiment was deployed using custom MATLAB scripts written using Psychophysics Toolbox (Brainard, 1997).

Functional localizer runs. Participants had to view a series of images against a black background and press a response button whenever an item was repeated. On each trial, 16 images were presented (0.8 s on, 0.2 s off), containing one repeat of an image that could occur at random. Trials were combined into blocks of 16 trials each containing either only objects or only scrambled objects. A single run of the functional localizers contained 12 such blocks (6 object blocks & 6 scrambled-object blocks). Stimuli in each block were chosen randomly from a larger pool of 80 naturalistic objects with the corresponding phase-scrambled objects (created by taking the 2D Fourier transform of each image, randomly shuffling the Fourier phase, and performing the Fourier inverse transform). This is a widely used method for functional localization of object-selective cortex. In practice, however, we observed no qualitative differences in our results upon using voxels activated during these functional localizer runs to further narrow down the voxels selected using anatomical masks. As a result, we did not use the functional localizer data, and all the analyses presented here are based on anatomical masks only.

Visual search task. In the present-absent search task, participants reported the presence or absence of an oddball target by pressing one of two buttons using their right hand. The response buttons were fixed for a given participant and counterbalanced across participants. Each search array had eight items, measuring 1.5° along the longer dimension, arranged in a 3 x 3 grid, with no central element to avoid fixation biases (as shown in Figure 2C). The entire search array measured 6.5°, with an average inter-item spacing of 2.5°. Item positions were jittered randomly on each trial according to a

uniform distribution with range ± 0.2°. Each trial lasted 4 s (1 s ON time and 3 s OFF time), and participants had to respond within 4 s. Each run had 64 unique searches (32 present, 32 absent) presented in random order, using the natural objects from Experiment 1. Target-present searches were chosen randomly from all possible searches such that all 32 images appeared equally often. Target-absent searches included all 32 objects. The location of the target in the target-present searches was chosen such that all eight locations were sampled equally often. In this manner, participants performed 8 such runs of 64 trials each.

Data acquisition. Participants viewed images projected on a screen through a mirror placed above their eyes. Functional MRI (fMRI) data were acquired using a 32-channel head coil on a 3T Skyra (Siemens, Mumbai, India) at the HealthCare Global Hospital, Bengaluru. Functional scans were performed using a T2*-weighted gradient-echo-planar imaging sequence with the following parameters: repetition time (TR) = 2s, echo time (TE) = 28 ms, flip angle = 79°, voxel size = 3 × 3 × 3 mm^3, field of view = 192 × 192 mm^2, and 33 axial-oblique slices for whole-brain coverage. Anatomical scans were performed using T1-weighted images with the following parameters: TR = 2.30 s, TE = 1.99 ms, flip angle = 9°, voxel size = 1 × 1 × 1 mm^3, field of view = 256 × 256 × 176 mm^3.

Data preprocessing. The raw fMRI data were preprocessed using Statistical Parametric Mapping (SPM) software (Version12; Welcome Center for Human Neuroimaging; https://www.fil.ion.ucl.ac.uk/spm/software /spm12/), running on MATLAB 2019b. Raw

images were realigned, slice-time corrected, co-registered to the anatomical image, segmented, and normalized to the Montreal Neurological Institute (MNI) 305 anatomical template. Repeating the key analyses with voxel activations estimated from individual subjects yielded qualitatively similar results. Smoothing was performed only on the functional localizer blocks using a Gaussian kernel with a full-width half-maximum of 5 mm. Default SPM parameters were used, and voxel size after normalization was kept at $3\times3\times3$ mm^3. The data were further processed using GLMdenoise (Kay et al., 2013). GLMdenoise improves the signal-to-noise ratio in the data by regressing out the noise estimated from task-unrelated voxels. The denoised time-series data were modeled using generalized linear modeling in SPM after removing low-frequency drift using a high-pass filter with a cutoff of 128 s. In the main experiment, the activity of each voxel was modeled using 83 regressors (68 stimuli + 1 fixation + 6 motion regressors + 8 runs). In the localizer block, each voxel was modeled using 14 regressors (6 stimuli + 1 fixation + 6 motion regressors + 1 run).

ROI definitions. The regions of interest (ROI) of Early Visual Cortex (EVC) and Lateral Occipital (LO) regions were defined using anatomical masks from the SPM anatomy toolbox (Eickhoff et al., 2005). All brain maps were visualized on the inflated brain using Freesurfer (https://surfer.nmr.mgh.harvard.edu/fswiki/).

Behavioral data analysis. Response times faster than 0.3 seconds or slower than 3 seconds were removed from the data. This step removed only 0.75% of the data and

improved the overall response time consistency, but did not qualitatively alter the results.

Model fitting for visual homogeneity.

We took the multi-dimensional embedding returned by the perceptual space experiment (Experiment 1) in 5 dimensions as the responses of 5 artificial neurons to the entire set of objects. For each target-present array, we calculated the neural response as the average of the responses elicited by these 5 neurons to the target and distractor items. Likewise, for target-absent search arrays, the neural response was simply the response elicited by these 5 neurons to the distractor item in the search array. To estimate the visual homogeneity of the target-present and target-absent search arrays, we calculated the distance of each of these arrays from a single point in the multidimensional representation. We then calculated the correlation between the visual homogeneity calculated relative to this point and the response times for the target-present and target-absent search arrays. The 5 coordinates of this center point was adjusted using constrained nonlinear optimization to maximize the difference between correlations with the target-present & target-absent response times respectively. This optimum center remained stable across many random starting points, and our results were qualitatively similar upon varying the number of embedding dimensions.

Additionally, we performed a leave-one-out cross-validation analysis to validate the number of dimensions or neurons used for the multidimensional scaling analysis in the visual homogeneity model fits. For each choice of number of dimensions, we

estimated the optimal centre for visual homogeneity calculations while leaving out all searches involving a single image. We then calculated the visual homogeneity for all the target-present and target-absent searches involving the left-out image. Compiling these predictions by leaving out all images by turn results in a leave-one-out predicted visual homogeneity, which we correlated with the target-present and target-absent response times. We found that the absolute sum of the correlations between visual homogeneity and present/absent reaction times increased monotonically from 1 to 5 neurons, remained at a steady level from 5 to 9 neurons and decreased beyond 9 neurons. Furthermore, the visual homogeneity using the optimal center is highly correlated for 5-9 neurons. We therefore selected 5 neurons or dimensions for reporting visual homogeneity computations.

Searchlight maps for mean activation (Figure 3A, Figure 5A)

To characterize the brain regions that encode visual homogeneity, we performed a whole-brain searchlight analysis. For each voxel, we took the voxels in a 3x3x3 neighborhood and calculated the mean activations across these voxels across all participants. To avoid overall activation level differences between target-present and target-absent searches, we z-scored the mean activations separately across target-present and target-absent searches. Similarly, we calculated the visual homogeneity model predictions from behaviour, and z-scored the visual homogeneity values for target-present and target-absent searches separately. We then calculated the correlation between the normalized mean activations and the normalized visual

homogeneity for each voxel, and displayed this as a colormap on the inflated MNI brain template in Figures 3A & 5A.

Note that the z-scoring of mean activations and visual homogeneity removes any artefactual correlation between mean activation and visual homogeneity arising simply due to overall level differences in mean activation or visual homogeneity itself, but does not alter the overall positive correlation between the visual homogeneity and mean activation across individual search displays.

Searchlight maps for neural and behavioural dissimilarity (Figure 3B, Figure 5B)

To characterize the brain regions that encode perceptual dissimilarity, we performed a whole-brain searchlight analysis. For each voxel, we took the voxel activations in a 3x3x3 neighborhood to target-absent displays as a proxy for the neural response to the single image. For each image pair, we calculated the pair-wise Euclidean distance between the 27-dimensional voxel activations evoked by the two images, and averaged this distance across participants to get a single average distance. For 32 target-absent displays in the experiment, taking all possible pairwise distances results in ${}^{32}C_2$ = 496 pairwise distances. Similarly, we obtained the same 496 pairwise perceptual dissimilarities between these items from the oddball detection task (Experiment 1). We then calculated the correlation between the mean neural dissimilarities at each voxel with perceptual dissimilarities, and displayed this as a colormap on the flattened MNI brain template in Figures 3B & 5B.

Experiment 3. Oddball detection for perceptual space (Symmetric/Asymmetric objects)

Participants. A total of 15 participants (11 males, 22.8 ± 4.3 years) participated in this experiment.

Paradigm. Participants performed an oddball visual search task. Participants completed 4032 correct trials ($^{64}C_2$ shape pairs x 2 repetitions) as two sessions in two days. We used a total of 64 baton shapes (32 symmetric and 32 asymmetric), and all shapes were presented against a black background. We created 32 unique parts with the vertical line as part of the contour. We created 32 symmetric by joining the part and its mirror-filled version, and 32 asymmetric objects were created by randomly pairing the left part and mirror flipped version of another left part. All parts were occurring equally likely. All other task details are the same as Experiment 1.

Experiment 4 Symmetry judgment task (fMRI & behavior)

Participants. A total of 15 subjects participated in this study. Participants had normal or corrected to normal vision. Participants had no history of neurological or psychiatric impairment. We excluded participants with metal implants or claustrophobia from the study.

Paradigm. Inside the scanner, participants performed two runs of one-back identity task (functional localizer), eight runs of symmetry judgment task (event-related design), and one anatomical scan. We excluded the data from one participant due to poor accuracy and long response times.

Symmetry Task. On each trial, participants had to report whether a briefly presented object was symmetric or not using a keypress. Objects measured 4° and were presented against a black background. Response keys were counterbalanced across participants. Each trial lasted 4 s, with the object displayed for 200 ms followed by a blank period of 3800 ms. Participants could respond at any time following appearance of the object, up to a time out of 4 s after which the next trial began. Each run had 64 unique conditions (32 symmetric and 32 asymmetric).

1-back task for functional localizers. Stimuli were presented as blocks, and participants reported repeated stimuli with a keypress. Each run had blocks of either silhouettes (asymmetric/symmetric), dot patterns (asymmetric/symmetric), combination of baton and dot patterns (asymmetric/symmetric) and natural objects (intact/scrambled).

Data Analysis

Noise Removal in RT. Very fast (< 100 ms) reaction times were removed. We also discarded all reaction times to an object if participant's accuracy was less than 80%. This process removed 3.6% of RT data.

Model fitting for visual homogeneity. We proceeded as before by embedding the oddball detection response times into multidimensional space with 3 dimensions. For each image, the visual homogeneity was defined as its distance from an optimum center. We then calculated the correlation between the visual homogeneity calculated relative to this optimum center and the response times for the target-present and target-absent search arrays separately. This optimum center was estimated using a constrained nonlinear optimization to maximize the difference between the correlations for asymmetric object response times and symmetric object response times. Other details were the same as in Experiment-2.

REFERENCES

Agrawal A, Hari K, Arun SP (2020) A compositional neural code in high-level visual cortex can explain jumbled word reading. Elife 9:e54846.

Arun SP (2012) Turning visual search time on its head. Vision Res 74:86–92.

Ashby FG, Maddox WT (2011) Human category learning 2.0. Ann N Y Acad Sci 1224:147–161.

Ayzenberg V, Kamps FS, Dilks DD, Lourenco SF (2022) Skeletal representations of shape in the human visual cortex. Neuropsychologia 164:108092.

Bainbridge WA, Dilks DD, Oliva A (2017) Memorability: A stimulus-driven perceptual neural signature distinctive from memory. Neuroimage 149:141–152.

Bertamini M, Makin ADJ (2014) Brain Activity in Response to Visual Symmetry. Symmetry (Basel) 6:975–996.

Bracci S, Op de Beeck H (2015) Dissociations and associations between shape and category representations in the two visual pathways. J Neurosci 36:432–444.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Carlin JD, Kriegeskorte N (2017) Adjudicating between face-coding models with individual-face fMRI responses. PLoS Comput Biol 13:e1005604.

Duncan J, Humphreys GW (1989) Visual search and stimulus similarity. Psychol Rev 96:433–458.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25:1325–1335.

Glezer LS, Riesenhuber M (2013) Individual variability in location impacts orthographic selectivity in the “visual word form area”. J Neurosci 33:11221–11226.

Goulet M-A (2020) Investigation of the Cognitive Mechanisms of Same and Different Judgments.

Haushofer J, Livingstone MS, Kanwisher N (2008) Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. PLoS Biol 6:e187.

Jacob G, Arun SP (2020) How the forest interacts with the trees: Multiscale shape integration explains global and local processing. J Vis 20:20.

Jaegle A, Mehrpour V, Mohsenzadeh Y, Meyer T, Oliva A, Rust N (2019) Population response magnitude variation in inferotemporal cortex predicts image memorability. Elife 8:e47596.

Katti H, Arun SP (2022) A separable neural code in monkey IT enables perfect CAPTCHA decoding. J Neurophysiol 127:869–884.

Kay KN, Rokem A, Winawer J, Dougherty RF, Wandell BA (2013) GLMdenoise: A fast, automated technique for denoising task-based fMRI data. Front Neurosci 7:1–15.

Kim J, Ricci M, Serre T (2018) Not-So-CLEVR: Learning same-different relations strains feedforward neural networks. Interface Focus 8.

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

Leopold D a, Bondar I V, Giese M a (2006) Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature 442:572–575.

Lukavský J, Děchtěrenko F (2017) Visual properties and memorising scenes: Effects of

image-space sparseness and uniformity. Attention, Perception, Psychophys 79:2044–2054.
Mohan K, Arun SP (2012) Similarity relations in visual search predict rapid visual categorization. J Vis 12:19–19.
Murdock BB (1960) The distinctiveness of stimuli. Psychol Rev 67:16–31.
Nickerson RS (1967) “Same”-"different" response times with multi-attribute stimulus differences. Percept Mot Skills 24:543–554.
Nickerson RS (1969) ‘Same’-‘different’ response times: A model and a preliminary test. Acta Psychol (Amst) 30:257–275.
Op de Beeck H, Wagemans J, Vogels R (2001) Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. Nat Neurosci 4:1244–1252.
Peelen M V., Kastner S (2014) Attention in the real world: Toward understanding its neural basis. Trends Cogn Sci 18:242–250.
Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, Kastman E, Lindeløv JK (2019) PsychoPy2: Experiments in behavior made easy. Behav Res Methods 51:195–203.
Petrov AA (2009) Symmetry-based methodology for decision-rule identification in same-different experiments. Psychon Bull Rev 16:1011–1025.
Pramod RT, Arun SP (2014) Features in visual search combine linearly. J Vis 14:1–20.
Pramod RT, Arun SP (2016) Object attributes combine additively in visual search. J Vis 16:8.
Pramod RT, Arun SP (2018) Symmetric Objects Become Special in Perception Because of Generic Computations in Neurons. Psychol Sci 29:95–109.
Proklova D, Kaiser D, Peelen M V (2016) Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate-Inanimate Distinction. J Cogn Neurosci 28:680–692.
Puebla G, Bowers JS (2022) Can deep convolutional neural networks support relational reasoning in the same-different task? J Vis 22:11.
Rhodes G, Jeffery L (2006) Adaptive norm-based coding of facial identity. Vision Res 46:2977–2987.
Ricci M, Cadène R, Serre T (2021) Same-different conceptualization: a machine vision perspective. Curr Opin Behav Sci 37:47–55.
Rust NC, Mehrpour V (2020) Understanding Image Memorability. Trends Cogn Sci 24:557–568.
Sani I, Stemmann H, Caron B, Bullock D, Stemmler T, Fahle M, Pestilli F, Freiwald WA (2021) The human endogenous attentional control network includes a ventro-temporal cortical node. Nat Commun 12:360.
Sasaki Y, Vanduffel W, Knutsen T, Tyler C, Tootell R (2005) Symmetry activates extrastriate visual cortex in human and nonhuman primates. Proc Natl Acad Sci U S A 102:3159–3163.
Serre T (2019) Deep Learning: The Good, the Bad, and the Ugly. Annu Rev Vis Sci 5:399–426.
Sripati AP, Olson CR (2010) Global Image Dissimilarity in Macaque Inferotemporal Cortex Predicts Human Visual Search Efficiency. J Neurosci 30:1258–1269.
Stewart N, Morin C (2007) Dissimilarity is used as evidence of category membership in multidimensional perceptual categorization: a test of the similarity-dissimilarity

generalized context model. Q J Exp Psychol (Hove) 60:1337–1346.
Storrs KR, Kietzmann TC, Walther A, Mehrer J, Kriegeskorte N (2021) Diverse Deep Neural Networks All Predict Human Inferior Temporal Cortex Well, After Training and Fitting. J Cogn Neurosci 33:2044–2064.
Sunder S, Arun SP (2016) Look before you seek: Preview adds a fixed benefit to all searches. J Vis 16:3.
Tartaglini AR, Feucht S, Lepori MA, Vong WK, Lovering C, Lake BM, Pavlick E (2023) Deep Neural Networks Can Learn Generalizable Same-Different Visual Relations.
Taylor KI, Devereux BJ, Acres K, Randall B, Tyler LK (2012) Contrasting effects of feature-based statistics on the categorisation and basic-level identification of visual objects. Cognition 122:363–374.
Thorat S, Proklova D, Peelen M V (2019) The nature of the animacy organization in human ventral temporal cortex. Elife 8.
Tyler LK, Chiu S, Zhuang J, Randall B, Devereux BJ, Wright P, Clarke A, Taylor KI (2013) Objects and categories: feature statistics and object processing in the ventral stream. J Cogn Neurosci 25:1723–1735.
Valentine T (1991) A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition. Q J Exp Psychol Sect A 43:161–204.
Valentine T, Bruce V (1986a) Recognizing familiar faces: The role of distinctiveness and familiarity. Can J Psychol Can Psychol 40:300–305.
Valentine T, Bruce V (1986b) The Effects of Distinctiveness in Recognising and Classifying Faces. Perception 15:525–535.
Van Meel C, Baeck A, Gillebert CR, Wagemans J, Op de Beeck HP (2019) The representation of symmetry in multi-voxel response patterns and functional connectivity throughout the ventral visual stream. Neuroimage.
Verghese P (2001) Visual search and attention: A signal detection theory approach. Neuron 31:523–535.
Wagemans J (1997) Characteristics and models of human symmetry detection. Trends Cogn Sci 1:346–352.
Weiner KS, Grill-Spector K (2012) The improbable simplicity of the fusiform face area. Trends Cogn Sci 16:251–254.
Wolfe JM (1998) What Can 1 Million Trials Tell Us About Visual Search? Psychol Sci 9:33–39.
Wolfe JM, Horowitz TS (2004) What attributes guide the deployment of visual attention and how do they do it? Nat Rev Neurosci 5:495–501.
Wolfe JM, Horowitz TS (2017) Five factors that guide attention in visual search. Nat Hum Behav 1:0058.
Zhivago KA, Arun SP (2014) Texture discriminability in monkey inferotemporal cortex predicts human texture perception. J Neurophysiol 112:2745–2755.
Zoccolan D, Cox DD, DiCarlo JJ (2005) Multiple Object Response Normalization in Monkey Inferotemporal Cortex. J Neurosci 25:8150–8164.

DATA AVAILABILITY

All data and code required to reproduce the results will be made available on OSF.

AUTHOR CONTRIBUTIONS

GJ and SPA designed the visual search experiments; GJ collected behavioral and fMRI data; GJ & SPA analyzed and interpreted the data; PRT & SPA designed the symmetry experiments; PRT collected fMRI data; GJ, PRT & SPA analyzed and interpreted the data; GJ and SPA wrote the manuscript with inputs from PRT.

ACKNOWLEDGEMENTS

We thank Divya Gulati for help with data collection of Experiments S4 & S5. This research was supported by the DBT/Wellcome Trust India Alliance Senior Fellowship awarded to SPA (Grant# IA/S/17/1/503081). GJ was supported by a Senior Research Fellowship from MHRD, Government of India.