Object Representations in Human Visual Cortex Formed Through Temporal Integration of Dynamic Partial Shape Views

Tanya Orlov and Ehud Zohary

Department of Neurobiology, The Alexander Silberman Institute of Life Sciences, and The Edmond and Lily Safra Center for Brain Sciences, Hebrew University, 91904 Jerusalem, Israel

We typically recognize visual objects using the spatial layout of their parts, which are present simultaneously on the retina. Therefore, shape extraction is based on integration of the relevant retinal information over space. The lateral occipital complex (LOC) can represent shape faithfully in such conditions. However, integration over time is sometimes required to determine object shape. To study shape extraction through temporal integration of successive partial shape views, we presented human participants (both men and women) with artificial shapes that moved behind a narrow vertical or horizontal slit. Only a tiny fraction of the shape was visible at any instant at the same retinal location. However, observers perceived a coherent whole shape instead of a jumbled pattern. Using fMRI and multivoxel pattern analysis, we searched for brain regions that encode temporally integrated shape identity. We further required that the representation of shape should be invariant to changes in the slit orientation. We show that slit-invariant shape information is most accurate in the LOC. Importantly, the slit-invariant shape representations matched the conventional whole-shape representations assessed during full-image runs. Moreover, when the same slit-dependent shape slivers were shuffled, thereby preventing their spatiotemporal integration, slit-invariant shape information was reduced dramatically. The slit-invariant representation of the various shapes also mirrored the structure of shape perceptual space as assessed by perceptual similarity judgment tests. Therefore, the LOC is likely to mediate temporal integration of slit-dependent shape views, generating a slit-invariant whole-shape percept. These findings provide strong evidence for a global encoding of shape in the LOC regardless of integration processes required to generate the shape percept.

Key words: anorthoscopic viewing; fMRI; lateral occipital complex; multivoxel pattern analysis; object shapes; temporal integration

Significance Statement

Visual objects are recognized through spatial integration of features available simultaneously on the retina. The lateral occipital complex (LOC) represents shape faithfully in such conditions even if the object is partially occluded. However, shape must sometimes be reconstructed over both space and time. Such is the case in anorthoscopic perception, when an object is moving behind a narrow slit. In this scenario, spatial information is limited at any moment so the whole-shape percept can only be inferred by integration of successive shape views over time. We find that LOC carries shape-specific information recovered using such temporal integration processes. The shape representation is invariant to slit orientation and is similar to that evoked by a fully viewed image. Existing models of object recognition lack such capabilities.

Introduction

The ability to identify objects despite great variation in their appearance is a fundamental characteristic of the visual system. Identity-preserving invariance to object transformations devel-

ops along the human ventral visual hierarchy and is most conspicuous in the lateral occipital complex (LOC) (Grill-Spector et al., 1999; Vuilleumier et al., 2002; Eger et al., 2008; Op de Beeck et al., 2008; Vinberg and Grill-Spector, 2008; Liu et al., 2009; Cichy et al., 2011). Classical models of object recognition typically start from a spatially extended retinal image of an object, in which the

Received May 14, 2017; revised Nov. 13, 2017; accepted Nov. 15, 2017.

Author contributions: T.O. and E.Z. designed research; T.O. performed research; T.O. analyzed data; T.O. and E.Z. wrote the paper.

This work was supported by a joint Edmond and Lily Safra Center for Brain Sciences (ELSC)/Ecole Polytechnique Fédérale de Lausanne (EPFL) Research Grant (E.Z.) and by the Israel Ministry of Absorption (T.O.). We thank Zvi Roth and Yuval Porat for helpful discussions.

The authors declare no competing financial interests.

Correspondence should be addressed to Tanya Orlov, Department of Neurobiology The Alexander Silberman Institute of Life Sciences The Hebrew University of Jerusalem Edmond J. Safra Campus, Jerusalem 91904, Israel. E-mail: tanya.orlov@mail.huji.ac.il.

DOI:10.1523/JNEUROSCI.1318-17.2017

Copyright © 2018 the authors 0270-6474/18/380659-20\$15.00/0

spatial layout of the object parts is available simultaneously. In such a case, local form elements can be integrated over space into higher-order units representing unique object shape (for review, see DiCarlo and Cox, 2007). Partial occlusion in stationary visual scenes can be a potential problem (Tang et al., 2014), but object recognition models can cope with it quite well (O'Reilly et al., 2013) by providing recurrent (top-down) signals to fill in the missing information. The models implement spatial integration rules, taking advantage of the fact that the viewed parts of the objects maintain a fixed relationship in retinotopic coordinates.

However, there are situations in which integration must be applied in the temporal domain. The anorthoscopic viewing paradigm provides such a case. Consider an object moving behind a narrow slit with only a small part of its contour visible at any moment. In these conditions, the translating object shape stimulates the same narrow retinal strip, thus creating retinotopic conflict between visual inputs presented in succession. However, the shape is typically perceived as an integrated whole despite the lack of the extended, simultaneous retinal image (Parks, 1965; Anstis and Atkinson, 1967). This perceptual phenomenon cannot be explained by the existing spatial integration models; if a simple spatial integration policy is applied, it would produce a jumbled pattern instead of a coherent shape.

To create such a coherent percept, the successive views of the stimulus must be integrated over time, taking into account the direction and speed of the occluded shape to piece together correctly its fragments. Theoretical and behavioral studies suggest that the fragments must be conveyed into a nonretinotopic space (Ögmen and Herzog, 2016) so that their spatial order could be recovered accurately (based on the motion estimation) to allow a spatially extended percept from retinotopically limited and conflicting visual inputs (Rock, 1981; Morgan et al., 1982; Casco and Morgan, 1984; Shimojo and Richards, 1986; Sohmiya and Sohmiya, 1994; Palmer et al., 2006; Rieger et al., 2007; Aydin et al., 2008; Palmer and Kellman, 2014).

Only a few fMRI studies have tried to assess the degree to which cortical areas are responsive to complex object shapes under similar slit-viewing conditions (Yin et al., 2002; Reichert et al., 2014). Slit-viewed drawings of familiar objects are perceived as a unified whole, but when they are distorted, they are seen as fragmented lines (Yin et al., 2002). Interestingly, both human LOC and the motion sensitive complex (MT+) were more active when the whole object was perceived than when the emergent percept was of multiple fragmented lines. However, to our knowledge, no study has assessed the degree to which cortical areas can encode shape identity under such conditions showing a preservation of shape identity encoding across slit orientations.

In the current fMRI study, we presented 3D shapes moving behind a narrow slit that could be either vertical or horizontal. Importantly, we used artificial, unfamiliar shapes to ensure that the integrated shape information is not contaminated by whole-shape priors. Crucially, we applied advanced information-based pattern analysis and focused on shape representations that preserve their shape selectivity regardless of the slit orientation. We reasoned that because in such conditions there is almost no overlap between the views seen through the vertical and horizontal slit at any moment, such representations are likely to truly encode temporally integrated whole shapes. We provide evidence that the LOC carries the richest and most accurate shape-specific information that is tolerant to changes in slit orientation.

Materials and Methods

Participants

Sixteen healthy volunteers (15 right-handed; mean age, 26.5 years; 4 females) participated in the fMRI study and completed all fMRI scans of our main experiment and behavioral tests. A subset of 11 volunteers (10 right-handed; mean age, 26.4 years; 2 females) also participated in a control fMRI experiment. The Helsinki Ethics Committee of Hadassah Hospital, Jerusalem, Israel approved the experimental procedure. Written informed consent was obtained from each participant before the procedure.

Stimuli and experimental setup

Visual stimuli were presented on an MR-compatible LCD screen (resolution: 1920 × 1800; refresh rate: 60 Hz; NNL LCD Monitor, Nordic NeuroLab) placed behind the scanner bore. The screen was made visible to the participants via a tilted mirror mounted on the head coil.

Main fMRI experiment

Participants. Sixteen participants completed the main experiment.

Stimuli. The stimuli were abstract, parametrically defined 3D shapes with complex shape characteristics (Fig. 1A) generated by a parametric shape model ("superformula"; Giell, 2003). Two model parameters were varied to create a 2D parameter space. Nine combinations of the parameters (i.e., nine points in the space arranged in a cross-like pattern) were chosen to create nine different shapes. The shape images were then equated (without background) in terms of mean luminance and contrast.

Experimental paradigm. We used an event-related fMRI paradigm. In slit runs, the slit orientation was either vertical or horizontal (Fig. 1B–D). The shapes were not shown before as a whole. In each trial, participants viewed one of the shapes moving behind a narrow slit in a fixed speed (6.6°/s) orthogonally to the slit orientation. The shape initially appeared in the slit after a 0.083 s delay, moved across the slit for 0.783 s until its edges disappeared, and then (after additional 0.15 s delay) reversed its direction and completed the move across the slit in another 0.783 s. The initial direction of the shape motion was chosen randomly. The active phase of the trial lasted 1.8 s followed by 1.2 s of "empty slit." Participants were required to fixate a fixation point. To maintain attention to the stimuli and to ensure shape perception, they were also asked to press a button whenever the same shape appeared in two consecutive trials (a one-back recognition task). The same shape was shown in 8.4% of such trials.

The slit width was set to 0.47° of visual angle. The shapes subtended 4.7 × 4.7° of visual angle. Therefore, only 1/10 of the shape contour was visible at any moment through the slit. The fixation point, a small red rectangle, was placed to the left/right of the vertical slit and on the top/bottom of the horizontal slit so that it would not interfere with the motion within the slit. Half of the participants fixated on the "left" and "top" fixation point in the vertical and horizontal runs and the rest fixated on the "right" and "bottom" fixation point. Each of the nine abstract shapes was presented three times over the course of either vertical or horizontal run in a counterbalanced manner using the Optseq procedure. The trials (3 s) were pseudorandomly embedded with fixation periods (an "empty slit" with a fixation point) ranging from 1 to 9 s. Each participant completed eight vertical-slit and eight horizontal-slit runs. The runs were acquired in an interleaved order (vertical, horizontal, vertical, horizontal, etc.). Each run was 115–119 s long (115–119 volumes) and ended with a 10–12 s fixation period. The total amount of trials in slit runs comprised 3 repetitions × 9 shapes × 16 runs = 432 trials.

Only after all slit runs were completed were subjects presented for the first time with full images of the shapes and performed the same task. In each trial of a full-image run (Fig. 1E), a full shape image was shown for 0.8 s followed by 1.2 s fixation period (a screen with a central small red rectangle). Each of the nine shapes was presented four times in each run. The total number of trials in full-image runs comprised 4 repetitions × 9 shapes × 8 runs = 288 trials. Eight full-image runs were completed in a counterbalanced manner. As above, the trials (2 s long) were pseudorandomly embedded with fixation periods ranging from 1 to 10 s. Each run lasted 113–120 s (113–120 volumes) ending with a 10–12 s fixation period. The total number of trials in full-image runs comprised 4 repetitions × 9 shapes × 8 runs = 288 trials.

Figure 1. Shape percept under slit viewing. **A**, 3D shapes were generated by a parametric shape model (“superformula”) with two model parameters. Nine combinations of the parameters were chosen to create the shapes. **B**, In each trial, a shape was seen moving behind a narrow vertical or horizontal slit (in a direction orthogonal to the slit orientation). Only 1/10 of the shape contour was visible at any moment. The shape moved behind the slit until it disappeared and then reversed its direction (i.e., it moved twice in a trial, in opposite directions). **C**, **D**, In fMRI trials, participants maintained fixation and pressed a button whenever the same shape appeared in two consecutive trials (one-back recognition task). Only a single trial is seen in the figure). Vertical- and horizontal-slit trials were presented in separate runs. Each participant completed eight runs of each type. **E**, After completion of all slit runs, subjects were presented with full images of the shapes and performed the same task in eight full-image runs.

Eye tracking. Participants were required to maintain their gaze on a fixation point throughout the experiment. To ensure that fixation was maintained in the slit fMRI trials, the monocular eye position of five (of 16) participants was tracked during three to six slit runs using a video-based, infrared illumination eye tracker (Eye Link1000; SR Research) with a sampling rate of 250 Hz. To determine fixation gaze stability, we computed spatial dispersion of eye position (Di Russo et al., 2003) across the active phase of a slit trial (when the shapes were moving). Because visual motion can skew gaze position (Zimmermann et al., 2012), the dispersion may be larger in the axis of motion direction (i.e., in X for vertical-slit trials and Y for horizontal-slit trials). The dispersion was small relative to the slit width (0.47°) ranging across participants between 0.16° and 0.43° (mean 0.26°) for the motion axis dimension and between 0.11° and 0.35° (mean 0.26°) for orthogonal dimension. This indicates that fixation was relatively stable in both cases. A 2-way ANOVA on the dispersion values with two factors slit orientation and dispersion dimen-

sion confirmed that spatial dispersion of eye position did not differ between the dimensions ($F_{(1,4)} = 0.21$, $p = 0.667$ and $F_{(1,4)} = 0.002$, $p = 0.970$, respectively).

Behavioral experiment

Participants. Sixteen subjects who participated in the main fMRI experiment also performed the complementary behavioral experiment.

Stimuli. The same shapes served as stimuli.

Experimental paradigm. Before and after the fMRI scans, participants performed a similarity-rating task under slit-viewing conditions. In the first rating session (before fMRI scans), the participants had not seen the whole shapes yet. The experiment started with a very short familiarization phase in which each of the nine shapes were shown once (translating behind a vertical slit). All participants reported that they clearly perceived a whole shape moving behind a slit. This ensured that they were able to perform the subsequent experiments (i.e., reliable similarity rating in the

Figure 2. Perceived shape dissimilarity under slit-viewing conditions before and after full-image exposure. Participants performed a similarity-rating task under slit-viewing conditions before (**A**) and after (**B**) full-image exposure. **A**, In each trial, the two shapes were presented sequentially through either a vertical or a horizontal slit (in separate blocks) and subjects reported their degree of similarity using a rating scale of 1–7. The similarity scores were averaged across the blocks. **B**, In the later test, the first shape in each pair was presented through a vertical slit and the second one translated behind a horizontal slit (or vice versa). **C, D**, Group mean perceptual distances between the shapes estimated based on the similarity-rating scores before (**C**) and after (**D**) the full-image exposure. Colors indicate the degree of dissimilarity for each shape pair. The distances between the shapes did not change substantially after the exposure.

behavioral experiment and a one-back shape recognition task in fMRI scans). Next, they were asked to rate shape similarity using a rating scale of 1–7, where 1 was the most dissimilar and 7 was the most similar. Subjects were instructed to use the entire scale for their responses and to rate shape similarity taking into account the 3D shape structure. In each trial, the two shapes were seen sequentially through a slit and then participants reported their degree of similarity by pressing a keyboard button. An example of such a trial is shown schematically in Figure 2A. The shapes moved exactly as in the fMRI trials: for 1.8 s each with interval of 1.4 s between them. Each trial ended after a response from the response button. Vertical- and horizontal-slit trials were presented in two different blocks to prevent generalization of shape information across slits. Each of the 45 shape pairs (36 pairs comprising images of different shapes and nine pairs with images of the same shape) was presented six times in a pseudorandom order across the blocks (three times per block). All other trial details were the same as in the fMRI main experiment.

In the second rating session, the participants had already seen the whole shapes in the course of the fMRI experiment. This time, to reinforce manifestation of global shape percept, the first shape in a pair was shown through a vertical slit and the second shape was presented behind a horizontal slit (and vice versa in other trial; Fig. 2B). All other experimental details were the same as in the first rating session. Each session lasted ~35 min.

Statistical analysis. First, we determined the reliability of the scores; that is, the extent to which our rating procedure yielded the same result for each shape pair on repeated trials in the same participant and across participants. We calculated both individual and intersubject reliability

with a split-half method and a Spearman–Brown correction (as commonly accepted in psychometrics; Carmines and Zeller, 1979). Specifically, we split the total number of pair assessment trials to two halves, averaged the scores per each half, and correlated the resulting two vectors of scores (using Pearson correlation). We then applied the Spearman–Brown formula [$2 * r / (1 + r)$] to yield the reliability estimate for the full set of trials (we performed the split 100 times to get a good estimate of reliability). The obtained reliability coefficients were further averaged across participants. To estimate the intersubject reliability, we split the total group of subjects in two halves and then proceeded as above. The within- and across-subject reliability was 0.94 and 0.98, respectively, indicating the high overall consistency of the similarity rating scores (reliability estimates vary within the range between 0 and 1).

In further analysis, the similarity-rating scores were averaged across trials per shape pair for each session in each participant. We normalized them to values between 0 and 1 and converted these values to perceived shape dissimilarity estimates (by subtracting the resulting values from 1). Each session's estimates for 36 pairs (comprising images of different shapes) were used, resulting in two dissimilarity vectors. Next, we assessed the degree of perceptual distance correspondence between the sessions (before and after seeing the full images) by calculating the Pearson correlation between the two dissimilarity vectors in each participant. The resulting r values (one per subject) were Fisher z -transformed to allow averaging across subjects. We then calculated the mean and SE and determined the lower and upper bounds of the 95% confidence interval. The mean and the bounds were transformed back into correlations (which are reported). Furthermore, to assess the degree of perceptual distance corre-

spondence within the group, we calculated the correlation between the dissimilarity vectors (pooled across the two behavioral sessions) across participants, resulting in the group mean and its confidence intervals. We also computed a pixelwise Euclidean distance measure of shape similarity (Grill-Spector et al., 1999). Perceptual distances between the shapes assessed by the similarity-rating task were tested for their correspondence to the pixelwise physical distances in each individual.

Control fMRI experiment

Participants. A subset of our participants ($n = 11$) also completed auxiliary control fMRI experiment.

Stimuli. The rationale behind this experiment was to ensure that shape recognition was not based on a distinct feature allowing for shape recognition without a need for a temporal integration process. To test this explicitly, we presented the same set of slit-dependent shape views in a random temporal order, thereby eliminating the possibility of temporal shape integration but preserving all features within a given frame. The same shapes served as stimuli at the same rate and during the same trial periods.

Experimental paradigm. The experiment included vertical and horizontal slit conditions only. However, the shape views (slivers) were now presented in a shuffled order, thereby eliminating the percept of both global shape and coherent motion (Morgan et al., 1982). The exact shuffle order was varied from trial to trial. To maintain attention to the stimuli and their shape features, participants performed a one-back recognition task: they were asked to press a button whenever the same set of shape views appeared in two consecutive trials. Each participant completed eight vertical-slit and eight horizontal-slit runs with the same parameters as above. The task was perceptually difficult, but it was performed well above chance [group mean d' (SD) = 1.0 (0.6)].

Auxiliary localizer scans

LOC localizer. All participants completed the LOC localizer scan. Grayscale images of real objects/abstract shapes and phase-scrambled versions of these images (while preserving the original power spectra of the images) were presented in a block design fashion. The localizer scan was composed of four conditions, objects, abstract shapes, phase-scrambled objects, and phase-scrambled abstract shapes, with 11 blocks for each condition. Each block lasted 8 s and was composed of 10 images (0.7 s per image with a 0.1 s interval of gray screen with a fixation point). The images were selected of a pool of 45 photographs for each category (real objects and abstract shapes) and were presented centrally, spanning $9 \times 9^\circ$. Participants were instructed to fixate on a central fixation point and to press a button whenever the same image appeared twice consecutively. This occurred on average once per block.

Early visual cortex mapping. We defined the early visual areas in all participants using two polar angle-mapping scans (Sereno et al., 1995; Engel et al., 1997). The polar angle stimuli consisted of a clockwise or counterclockwise rotating wedge composed of monochromatic checkerboard patterns with counter-phasing flicker frequency of 7.5 Hz. The radial size of the pattern segments was adjusted to approximate the cortical magnification factor (Sereno et al., 1995). Each wedge covered 45° of arc and extended from the fixation point to 8.5° into the visual periphery. Each scan, either clockwise or counterclockwise, comprised six cycles of the rotating wedge of 24 s duration. Participants were instructed to fixate on a central fixation point. To ensure that fixation was maintained throughout the run, the color of the fixation point changed for 0.1 s 9 times per scan, and participants were asked to indicate the change with a button press.

Visual motion localizer. All participants completed the MT+ localizer scan (Huk et al., 2002). Participants were presented with a dot pattern that was either moving at $8^\circ/\text{s}$ alternating direction radially inward and outward from fixation once per second (10 s) or stationary (10 s). The task was to fixate on a central fixation point and covertly count the number of fixation point blinks (presented 25 times throughout the scan). During a motion block, 1200 white dots on a black background were presented within a $16^\circ \times 16^\circ$ aperture centered at the fixation point. Each dot appeared for 0.1 s and was then replaced by another dot at a randomly selected position. The moving/stationary pair of blocks was repeated 12 times.

Body/face localizer. A subset of our participants ($n = 11$) also completed the body/face localizer scan. This included three experimental conditions (headless bodies, faces, and everyday objects) with nine blocks for each condition. Each block lasted 10 s and was composed of 12 images (0.7 s per image with a 0.13 s interval). The three conditions were counterbalanced and interleaved with “baseline” blocks, including a fixation point only. The images were selected of a pool of 58 photographs for each category and were presented centrally, spanning $9 \times 9^\circ$. Participants performed the same one-back task as above.

MRI data acquisition and processing

The blood oxygenation level-dependent (BOLD) fMRI measurements were acquired using a 3 tesla Magnetom Skyra Siemens scanner and a 32 channel head coil. The fMRI protocols were based on a multiband echoplanar imaging sequence with the following parameters: TR = 1 s, TE = 30 ms, flip angle = 62° , acquisition matrix = 64×64 , FOV = 192×192 mm, multiband factor = 3. A total of 42 slices with 3 mm slice thickness (with no gap) were oriented in an oblique position covering the whole brain, with functional voxels of $3 \times 3 \times 3$ mm. In addition, high-resolution T1-weighted magnetization-prepared rapid acquisition gradient-echo images were acquired ($1 \times 1 \times 1$ mm resolution).

MRI data were processed using BrainVoyager QX software (Version 2.8; Brain Innovation). Subsequent analyses were performed using MATLAB (version 7.11; The MathWorks). Head motion correction and high-pass temporal filtering in the frequency domain (two cycles/total scan time) were applied. The slice-based functional images were coregistered with the high-resolution 3D anatomical image. The complete functional datasets were then resampled into a standard 3D space (Talairach and Tournoux, 1988) with 3 mm isotropic resolution. All functional data were analyzed with no spatial smoothing. The cortical surface of each participant was reconstructed from the high-resolution T1-weighted scan, which was transformed into the standard brain template. To assess BOLD responses, we applied a conventional general linear model (GLM). Repressors were obtained by convolving stimulus presentation with a double-gamma hemodynamic response function.

Selection of regions of interest (ROIs)

Standard ROI set. Eleven bilateral ROIs were defined for each of the 16 participants. These were volume-based ROIs with 3 mm isotropic voxel resolution restricted to the gray matter. Figure 3 shows the ROIs of one representative participant. Six visual retinotopic areas (V1, V2, V3, V3ab, V4, and LO-1) were identified based on retinotopic mapping. Using a phase-encoding approach (Sereno et al., 1995; Engel et al., 1997), we constructed individual polar maps and overlaid them on inflated cortical surfaces (as implemented by BrainVoyager QX software). The retinotopic borders were delineated manually (Wandell et al., 2007) in each hemisphere. The volume-based ROIs for V1, V2, V3, V3ab, hV4, and LO-1 in both hemispheres were then extracted comprising 394 (95) [group mean (SD)], 309 (90), 250 (46), 241 (65), 97 (29), and 93 (43) voxels, respectively. MT+ was defined based on its greater BOLD response to moving dots compared with stationary dots. A false discovery rate (FDR) criterion was applied to correct for multiple comparisons. MT+ comprised significant contiguous voxels [$q(\text{FDR}) < 0.05$] located in the posterior inferior temporal sulcus (Huk et al., 2002). The group mean ROI size for MT+ was 164 (47) voxels across both hemispheres.

Object-selective regions were identified using our LOC localizer (Malach et al., 1995). To better separate between LOC and object-selective foci in the dorsal visual pathway (as well as between different LOC subparts), we first created three bilateral anatomical masks matching the standard anatomical boundaries for these regions (Grill-Spector et al., 1998; Golarai et al., 2007; Vinberg and Grill-Spector, 2008; Julian et al., 2012). The masks included the following: (1) the lateral occipital cortex, (2) the posterior and midfusiform gyrus, and (3) the caudal part of the intraparietal sulcus (cIPS). They did not include retinotopic cortex and MT+. LOC was identified within the lateral occipital cortex and the fusiform gyrus in both hemispheres (i.e., within both first and second anatomical masks) using the standard contrast: intact images > their scrambled counterparts, with [$q(\text{FDR}) < 0.05$] (Grill-Spector et al., 1999). To avoid issues related to the variability of LOC sizes across participants, we lim-

Figure 3. Main set of ROIs. Individual-participant ROIs are shown on the corresponding cortical surface from a posterior view. **A**, Polar map of the participant. White solid lines indicate the borders of six visual retinotopic areas (V1, V2, V3, V3ab, V4, and LO-1). Colors reflect the preferred polar angle. **B**, MT+ (marine), LOC (red), and cIPS (green) ROIs. **C**, In addition, two LOC subregions were defined separately: LO (dark red) and pFs (dark magenta).

ited the number of voxels in the LOC ROIs: we sorted the voxels according to their t -values (in the standard LOC contrast) and picked the 350 voxels with highest values (if an individual LOC comprised <350 voxels, all of these voxels were taken for the analysis). This number was chosen conservatively based on LOC sizes (in cubic millimeters) reported previously in the literature (Golarai et al., 2007; Vinberg and Grill-Spector, 2008; Julian et al., 2012). LOC voxels were selected based on their t -values regardless of their contiguity (as long as they were within the standard anatomical boundaries); in most cases, they were contiguous.

In addition, using the same contrast (intact images > scrambled images), we identified two LOC subregions: the lateral occipital (LO) and posterior fusiform (pFs) regions (Grill-Spector et al., 1998; Vinberg and Grill-Spector, 2008). For each participant, we first selected significant voxels [$q(\text{FDR}) < 0.05$] located within the corresponding anatomical mask (either first or second). Then, the number of voxels was limited to the 200 or 150 most responsive voxels in LO or pFs, respectively, taking into account their reported sizes (Vinberg and Grill-Spector, 2008). Finally, object-selective regions in the dorsal visual pathway were also identified (using the third mask) and included cIPS and sometimes the adjacent transverse occipital sulcus (Grill-Spector et al., 1998, 2000).

Because LOC overlaps with category-specific visual regions (Downing et al., 2007), we defined a set of additional bilateral ROIs in 11 of the 16 participants using our body/face localizer. These included extrastriate body area (EBA) and fusiform body area (FBA) (Downing et al., 2001), as well as fusiform face area (FFA) and occipital face area (OFA) (Kanwisher et al., 1997). EBA and FBA were identified as lateral and ventral occipital regions, respectively, that were activated more strongly to images of human bodies than to objects [$q(\text{FDR}) < 0.05$]. FFA and OFA were defined in their standard anatomical locations based on their greater BOLD response to faces compared with objects [$q(\text{FDR}) < 0.05$; OFA was identified in 10 participants]. Voxels overlapping with the previously delineated retinotopic cortex and MT+ were excluded. The group mean sizes for the initially defined EBA, FBA, FFA, and OFA were 192 (31), 48 (30), 79 (30), and 51 (25) voxels, respectively. Six new ROIs were then constructed. First, in each participant, we labeled all mutual voxels between the LOC ROI and each body/face ROI. These voxels (comprising on average 29% from the total LOC volume) were grouped into “shared LOC” ROIs. The rest of LOC voxels represented “nonshared LOC” ROIs. EBA, FBA, FFA, and OFA ROIs were eventually restricted to the voxels that do not overlap with LOC.

Fixed-size ROI set: equalizing the number of voxels in each ROI. Because ROI size can affect the results of multivariate pattern analysis in the ROIs (Schreiber and Kretzelberg, 2013; Walther et al., 2016), our fixed-size ROI set was created so that all areas would have an equal number of voxels. As above, we applied activation-based voxel ranking, but this time, we chose voxels that responded best to all slit-viewed shapes regardless of slit orientation. Before the procedure, V4 and LO-1 were combined (due to their small sizes in some participants). V2 and V3 voxels were also pooled into one ROI. Therefore, the new set comprised nine ROIs in each of our

16 participants: V1, V2 + V3, V3ab, V4 + LO-1, MT+, cIPS, LOC, LO, and pFs. The ranking procedure was applied iteratively. In each iteration, ROI data from one vertical-slit run and one horizontal-slit run were taken jointly. Then, a single t -value for each voxel was calculated in the contrast testing all shape conditions against all rest periods (regardless of a slit orientation). Next, we sorted the voxels according to their t -values and took the 100 most active voxels that responded best to all slit-viewed shapes (a few individual ROIs comprised <100 voxels; all of these voxels were taken for the analysis). The remaining data were used for multivoxel pattern analysis (MVPA). The procedure was repeated eight times (according to the number of slit runs): the first vertical and first horizontal run were taken for voxel selection in the first iteration, etc. Therefore, in each iteration, an independent dataset was used for voxel selection. Importantly, MVPA and feature selection were always performed on different datasets to exclude issues of circular inference. The MVPA results were averaged across the iterations.

Note that, in a subset of our participants ($n = 11$), six additional ROIs were defined: “shared LOC”, “nonshared LOC”, EBA, FBA, FFA, and OFA. The EBA, FBA, FFA, and OFA ROIs were pooled into one body/face ROI (due to their small sizes in some participants). Using the above voxels’ ranking procedure, “shared LOC” and “nonshared LOC” ROIs, as well as combined body/face ROI, were similarly limited to 100 voxels that responded best to all slit-viewed shapes (regardless of slit orientation). Therefore, the new set comprised 12 ROIs in these participants.

ROI-based MVPA

We used support vector machine (SVM) classification (based on LIBSVM; Chang and Lin, 2011) and a complementary correlation analysis (Haxby et al., 2001) for MVPA. Although these methods probe somewhat different aspects of the brain activation patterns, we expected their results to largely correspond to each other (Walther et al., 2016).

Cross-validation scheme. We applied a cross-validation approach to assess the degree to which the voxel patterns of activation convey shape information in the various slit conditions. To that end, we first partitioned our data. In the main ROI analysis (for the standard ROI set), all collected data samples were used for MVPA (comprising 3 viewing conditions $\times$ 8 runs = 24 runs in total). For classification analysis, each viewing condition dataset was split into two unequal, nonoverlapping subsets composed of six and two runs. The data from six runs were used to train a SVM model. The remaining data from the other two runs were used to test the SVM performance. The splitting process was performed multiple times to reduce the variance of the performance (using all possible permutations of a split of the eight runs into the two subsamples, $n = 28$).

In the correlation analysis, we divided the data using the classical split-half approach (Haxby et al., 2001; Chan et al., 2010) and the unequal-split approach as in our SVM analysis (see also Walther et al., 2016). Both split schemes gave similar correlation-based decoding results. We used here the unequal split to be consistent with the classification analysis so that the variance of both SVM and correlation results could be similarly

attenuated by cross-validation (due to the same cross-validation scheme). The raw data from runs within each subgroup (of six and two runs) were then z -transformed and concatenated.

In our equal-size ROI analysis, we divided the raw data using two iteration loops. In the “outer” loop, one vertical-slit and one horizontal-slit run was chosen for voxel selection (and therefore was excluded from MVPA). This was done iteratively using the eight possible choices of the one run (e.g., first vertical and first horizontal run were chosen in the first iteration, etc., as explained previously). The remaining seven runs were further divided to two subgroups for MVPA using the “inner” loop. In classification analysis, five runs were used for training the SVM model and the two other runs were used for cross-validation testing. In the correlation analysis, the raw data from runs within each subgroup (a block of five or two runs) were z -transformed and concatenated. The inner loop cross-validation procedure was repeated 21 times (thus including all possible partitions of seven runs into subgroups of five and two runs).

Shape discrimination analysis. We had nine predictors (i.e., one per each shape image) to fit the data in each voxel. In the classification analysis, this was done separately for each run. In the correlation analysis, the predictors were computed per run block consisting of either six or two runs in the main ROI analysis (using the standard ROI set) or either five or two runs in the equal-size ROI analysis (using the fixed-size ROI set). We then applied a standard GLM to estimate the BOLD response per each shape in each separate run (in the classification analysis) or subgroup (in the correlation analysis). Before analyzing the resulting patterns of regression coefficients, we used a denoising procedure designed to account for the noise in the patterns of activation. This was done because noise normalization enhances the reliability of both pattern classification and the estimates of dissimilarity between the activation patterns (Walther et al., 2016). The noise structure in each voxel within a ROI was assessed from the residual error of the GLM. We then normalized each voxel's regression coefficient by the SD of its residual, thus suppressing the contribution of noisy voxels that can have high (either positive or negative) regression coefficients due to high noise (Misaki et al., 2010). In addition, we accounted for the noise covariance between the voxels in each ROI (for further details, see Walther et al., 2016).

To perform pattern classification, the regression coefficients were normalized by subtracting from each coefficient the population mean and dividing by the population SD. The SVM model used a linear kernel function and one-versus-rest multiclass classification scheme (Reddy et al., 2010). SVMs were trained iteratively (based on the above cross-validation scheme) to distinguish between the shapes in one viewing condition (e.g., vertical-slit) and its generalization performance was tested in the other two conditions (horizontal-slit or full-image runs and vice versa). In total, there were six combinations of viewing conditions. Classification performance across different slit-viewing conditions (two combinations) and across slit and full viewing conditions (four combinations) was averaged across the relevant combinations. Shape classification was also tested in full-image runs (to get a benchmark for performance) with generalization performance assessed across different runs.

Consequently, $3 \times 9 \times 9$ confusion matrices were calculated by assessing the proportion of trials of each predicted shape (by the classifier) for each input shape: each row in the matrices represents a specific shape that was shown and each column represents a predicted shape. Therefore, each entry in the matrix $P(i,j)$ is simply the proportion of cases in which the classifier classified shape i as being shape j . Correct classification accuracy is represented by the diagonal elements $P(i,i)$ and was averaged to receive a mean value across all shapes. The accuracy was further averaged across the cross-validation iterations in each individual ROI. Chance level decoding is $1/9 = 11\%$. The original confusion matrices were used for further analysis (see next section).

In the correlation analysis, the response vectors per image were first extracted per iteration for each run subset and normalized by subtracting the voxel's mean activation level across shape-image conditions. In total, 18 response vectors ($9 \text{ shapes} \times 2 \text{ subsets of the data}$) were extracted for each viewing condition (i.e., for vertical-slit, horizontal-slit, and full-image viewing). Next, for each pair of images, we calculated the Pearson correlation between their response patterns from different viewing con-

ditions. That is, response vectors from the subset #1 (in one viewing condition) were cross-correlated with the response vectors from the subset #2 (in another viewing condition), resulting in 6 versions of 9×9 correlation matrix. Correlation values were Fisher's z -transformed. The two across-slit correlation matrices and four slit- versus full-viewing matrices were averaged. The benchmark correlation matrix was also calculated for the response vector pairs in the full-viewing condition across independent subsets of these vectors. Therefore, the end result was $3 \times 9 \times 9$ correlation matrices: 2 of them measured the degree of similarity of the response vectors across different viewing conditions and 1 across different runs in the same full-viewing condition. To assess the degree of shape discrimination, we subtracted the mean off-diagonal scores (i.e., correlations between different-shape patterns) from the mean diagonal scores (i.e., correlations between same-shape patterns). The resulting excess correlation was averaged across the iterations in each ROI to obtain a shape discrimination score per participant. The original correlation matrices were used for further analysis (see next section).

Shape-related representational similarity analysis. To assess the degree of correspondence between the similarity of the evoked brain responses to a pair of images and their perceptual resemblance (as reflected by their similarity rating score), we used representational similarity analysis (RSA). Perceptual distances between each two shapes (i.e., dissimilarity measures) were estimated in two separate behavioral sessions performed before and after scanning (as explained in “Behavioral experiment” section). The resulting estimates were highly correlated across the sessions (see Results). Therefore, the distances were pooled across the two sessions for each participant. Similarly, the distances were highly correlated across participants (see Results), so we decided to use their group mean average for the RSA.

To assess the degree of dissimilarity between brain activation patterns for each pair of shapes, the off-diagonal correlation scores in the correlation matrices (from the above discrimination analysis) were converted into dissimilarity measures. The dissimilarity (distance) was calculated as $1 - \text{Pearson's } r$, yielding the representational dissimilarity matrix (RDM) (Kriegeskorte et al., 2008; Nili et al., 2014). Misclassifications (off-diagonal scores) in the classification matrices also provide information about the distance between patterns corresponding to different shapes (Haxby et al., 2014; Walther et al., 2016). We normalized the resulting values in the matrix (by dividing each entry by the maximum value in the matrix) and subtracted the resulting values from 1 (to get a pattern dissimilarity measure). Because our distance measures are derived after multivariate noise normalization and cross-validated, they are (relatively) independent of the patterns' noise and therefore reflect the true distances between the patterns (Walther et al., 2016).

Finally, we computed the Pearson correlation between the shape perceptual distances and response patterns' RDMs (Ejaz et al., 2015) per ROI. We then squared the resulting correlation coefficient to assess the proportion of variance in the RDMs (R^2) that can be accounted for by the perceived shape dissimilarity.

Whole-brain searchlight MVPA

We applied a volume-based searchlight analysis (Kriegeskorte et al., 2006). For each participant, we searched iteratively through the brain using a cubic search window comprising 125 voxels ($5 \times 5 \times 5$; $15 \times 15 \times 15 \text{ mm}^3$, respectively) for a multivoxel pattern that carries slit-invariant shape-related information. On each iteration, the window was centered on a new Talairach voxel and we performed cross-validated classification analysis in the exact same way as for the standard ROI set. A classifier was trained to discriminate between shapes under one slit orientation and its generalization performance was tested under the other orientation. We calculated classification performance per cluster (centered on a specific Talairach voxel), extracted the cluster's activation pattern RDM and assessed the proportion of variance in that RDM accounted for by the perceived shape dissimilarity.

Experimental design and statistical analysis

All experiments were planned as within-subjects designs. Statistical analyses of shape perceptual similarity judgments and fMRI-based estimates of shape information were performed on datasets from 16 human par-

Figure 4. Confusion matrices for shape decoding in V1 and LOC across viewing conditions. **A–C**, Top, A classifier was trained to distinguish between shapes in one viewing condition and its generalization performance was tested in the other conditions. Specifically, the training and testing was performed across different slit conditions (**A**) and across slit and full viewing conditions (**B**). **C**, Shape classification was also tested across different runs of the same full-image condition. Middle, Insets, Proportion of the shape image that was shared on the retina across different training and testing conditions. This indicates how much shape information could have been generalized merely due to the shared retinal image without shape integration over time. Bottom, Group mean confusion matrices for LOC and V1 in different generalization tests. Rows are the actual shapes; columns are the predicted shapes. Main diagonal values are the correct classifications; off-diagonal values are misclassifications. Chance level decoding is classification performance at 11%. Whereas both V1 and LOC show above-chance performance for the full viewing condition (**C**), only LOC shows above-chance generalization across slit orientations (**A**) and across slit and full viewing conditions (**B**).

ticipants (four females). Statistical analysis of behavioral data is detailed in the “Behavioral experiment” section.

ROI-based statistical analysis of shape information. Statistical analysis was applied to the shape discrimination scores [i.e., classification accuracy (%) and excess correlation] and R^2 (i.e., the explained variance in the classification-based and correlation-based RDMs). In each ROI, for each generalization rule (e.g., generalization across different slit conditions), we determined whether the accuracy (across participants) was significantly greater than chance level using one-sample (two-tailed) t tests. The same approach was used to determine whether excess correlation and R^2 were significantly greater than zero. To differentiate between the various regions of interest (i.e., those showing view-invariant shape information vs those that do not have this capability) we applied a two-way repeated-measures ANOVA: The factors were ROI identity and the generalization capability. The generalization factor was a three-level factor that reflects the type of generalization: (1) across different slit-viewing conditions (2) across slit and full-viewing conditions, and (3) across different runs of the same full-viewing condition (Fig. 4, inset). The number of levels in the ROI identity factor was equal to the total number of ROIs (in either the standard or fixed-size ROI set), excluding LO and pFs (because the majority of LO and pFs voxels are contained within LOC ROI). Having found a significant interaction term using this analysis, we

continued with a subsequent analysis of simple main effects using one-way ANOVA separately in each ROI. We also used *post hoc* paired (two-tailed) t tests between ROIs to test whether the shape information estimate (e.g., classification accuracy) was different between ROIs. Statistics were performed based on all participants in the study.

In the control fMRI experiment, we tested whether slit-invariant shape discrimination and R^2 change when the same set of slit-dependent shape views (slivers) are presented in a shuffled temporal order. A repeated-measures ANOVA with two factors, ROI identity (levels indicating different ROIs) and temporal order (two levels, indicating the correct or random temporal order) was run, followed up with *post hoc* t tests. Statistics were performed based on a subset of our participants ($n = 11$).

The simple main effects and t test statistics were corrected for multiple comparisons using Sidák’s correction. In addition, to corroborate the robustness of the perceptual–neural correspondence estimate (R^2), we ran a permutation analysis. The values within each previously calculated confusion matrix were permuted across the matrix space for each ROI in each participant. This procedure was repeated 2000 times and the same RSA analysis as described in the previous paragraph was repeated, resulting in a chance distribution of the group mean R^2 in each ROI. Given these chance distributions, we assessed the significance of the estimates. The obtained values were very close to the

values from the parametric tests. We therefore reported the results of the parametric tests alone.

Whole-brain statistical analysis. Statistical analysis was applied to cross-slit classification accuracy (percentage) per searchlight cluster and the explained variance in the cluster's classification-based RDM (R^2). Clusterwise one-sample two-tailed t tests were used to identify search windows showing significant values of the estimates across 16 participants. The resulting t -value statistical maps were corrected for multiple comparisons [$q(\text{FDR}) < 0.001$]. For comparison, we present the location of the most significant clusters in the searchlight analysis and the location of the most significant ROI showing the greatest cross-slit shape sensitivity (LOC) (Fig. 10). To that end, we calculated the group probability map for LOC voxels. For this purpose, all individual LOC ROIs (from the standard ROI set) were combined into one ROI. The probability of each Talairach voxel was assessed by counting the proportion of individuals in which that specific voxel was included in the individual-based ROI.

Results

We used nine novel artificial 3D shapes (which were not seen previously) to maximize the need for temporal shape integration to recognize the image (Fig. 1A). In each fMRI trial, 16 participants viewed one of the shapes moving behind a narrow vertical or horizontal slit so that only 1/10 of its contour was visible at any instant. As Figure 1B demonstrates, these individual views of the shape fragments were poor indicators of the overall shape when seen alone, but the participants recognized the overall shapes using temporal integration exceedingly well. To ensure that they paid attention to the overall shape during the scan, they performed a one-back shape-matching task [group mean d' (SD) = 2.8 (0.7)], pressing a button whenever the same whole shape appeared in two consecutive trials. After all slit runs were acquired (Fig. 1C,D), full images of the shapes were shown in separate runs (in the context of the same task; Fig. 1E).

A subset of our participants ($n = 11$) completed additional (control) slit runs in which the slit-limited slivers were presented in a shuffled order, thereby eliminating the possibility of temporal integration for shape recovery. The rationale for this auxiliary experiment was to rule out the possibility that encoding of shape information was based on specific features (e.g., cues) that may have been present in the individual slit-seen slivers. All participants reported that they perceived a (static) mixture of vertical or horizontal shape fragments presented in a jumbled sequence and failed to see a whole shape.

Behavioral experiment: slit-viewed shape perception before and after full shape exposure

How accurate are the shape percepts established under slit viewing? If they truly represent the whole, integrated shape, then the perceptual distances between the shapes should not change substantially after full shape exposure. To test this hypothesis, we measured perceptual shape similarity twice, before and after full shape presentation during the course of full-image fMRI scans. In the first behavioral session (that preceded all fMRI scans including the full-image scans), the same 16 participants saw pairs of sequentially presented shapes through either vertical or horizontal slit (in two different blocks) and reported their degree of similarity using a rating scale of 1–7 (Fig. 2A). Figure 2C represents the resulting group mean shape dissimilarity (obtained by averaging of rating scores per shape pair across blocks and participants, normalizing them to values between 0 and 1 and subtracting the resulting values from 1). When the participants were tested for the second time, they had already seen full shape images during the course of full-image fMRI scans. In this case, to reinforce manifestation of global shape percept, the first shape in each pair was presented through a vertical slit and the second one moved behind a horizontal

slit and vice versa (Fig. 2B). Corresponding group mean shape dissimilarity matrix is shown in Figure 2D.

Perceptual shape dissimilarities acquired in the two sessions were highly correlated between the sessions in each individual [group mean Pearson's $r = 0.868$ (95% confidence interval, 0.842 – 0.890)]. This indicates that full shape exposure did not add a lot: the percept of the whole shape in slit-viewing conditions was robust even before the observers saw the full shape images. Furthermore, the judged perceptual distances between the shapes (pooled across the tests) were also highly correlated across participants [group mean Pearson's $r = 0.820$ (0.799 – 0.839)]. This across-subject agreement indicates that the structure of shape perceptual space exhibits a common configuration across individuals.

It is possible that the participants might have performed the similarity-rating task (and the one-back shape recognition task in fMRI scans) without shape integration. However, we believe that this possibility is unlikely. First, all of our subjects reported that they could see the whole shape in the initial familiarization trials preceding all of our experiments (see Materials and Methods). In addition, the fact that the cross-slit similarity judgments (Fig. 2B,D) were very similar to the similarity judgments in the fixed-slit configuration (Fig. 2A,C) provides further evidence that the participants were assessing global shape (resulting from temporal integration) rather than a local 2D spatial feature match. Finally, we computed a pixelwise Euclidean distance measure of shape similarity (see Materials and Methods). This distance estimate indicates how the shapes are similar to each other physically (i.e., in their spatially extended retinal images) when the layout of the shape parts is available simultaneously. Perceptual distances between the shapes assessed during slit viewing were highly correlated with the pixelwise physical distances in each individual [group mean Pearson's $r = 0.773$ (0.735 – 0.806)]. It seems unlikely that the correlation would be significant if the observers rated the shapes on the basis of specific slit-dependent shape features without having a percept of the whole, spatially extended shape.

ROI-based MVPA

Our main goal was to identify brain regions that encode temporally integrated shape identities. For that purpose, we studied the multivoxel patterns of fMRI activity evoked by the shapes under slit-limited and full viewing in three groups of ROIs (Fig. 3): early visual cortex (V1, V2, and V3), dorsal stream areas (V3ab, MT+, and object-selective cIPS), and ventral stream areas (V4, LO-1, and object-selective LOC). In addition, we isolated two functionally distinct subregions within LOC, a posterior lateral subregion (area LO) and a more ventral (and anterior) subregion (pFs), and defined them separately (Grill-Spector et al., 1999; Op de Beeck et al., 2008; Vernon et al., 2016).

We hypothesized that, if a certain ROI represents the integrated shape percept, then it should carry information about the shape regardless of slit orientation. Furthermore, if temporally integrated and conventional shape representations converge on the same neural network, then the ROI should also generalize shape-specific information across slit-viewing and full-viewing conditions. To test our hypothesis, we trained a classifier to distinguish between the activation patterns in one viewing condition (e.g., vertical slit) and tested its generalization performance in the other conditions (horizontal-slit or full-image runs and vice versa; see Materials and Methods and Fig. 4, top). For comparison, shape classification was also tested across different full-image runs.

Because subjects kept a stable fixation during the experiment (see Materials and Methods for details about eye tracking), each

shape stimulated the same narrow strip of the retina over time, either vertical or horizontal. The slit width was set to be equal to 1/10 of the shape image. Therefore, at any time point, only 10% of the image was common to both the slit-viewing condition and the full-viewing condition. In fact, strictly speaking, the same retinal image was available only at one time point (per motion direction, when the central 10% of shape was visible through the slit). Vertical-slit and horizontal-slit runs could share only 1% of the retinal images (without taking into account the retinal magnification factor), whereas training and testing across different full-image runs meant that the full retinal image was available (Fig. 4, middle, insets). This gives a proxy estimate to how much information about the shape may be generalized across these conditions due to shared retinal image without any shape integration over time.

Figure 4 (bottom) shows the group mean confusion matrices for shape decoding across viewing conditions in V1 and LOC. Values along the main diagonal indicate correct classifications and the off-diagonal values indicate misclassifications by the decoder. Not surprisingly, both V1 and LOC represented the shapes well above chance when full shape images were presented (Fig. 4C). In this case, the whole shape was mapped onto the retina and this spatially extended retinal image was the same in both training and testing trials (i.e., 100% of the shape image was shared). In contrast, only LOC showed above-chance generalization performance across slit orientation (Fig. 4A). The tiny degree of joint retinal image information in this case (1%) is certainly not enough to decode shapes across different slit conditions. Clearly, only a slit-invariant representation of the integrated shape can support this generalization. Similarly, LOC, but not V1, showed robust generalization across presentation mode (slit vs full-viewing; Fig. 4B). This suggests that the temporally integrated and conventional shape representations have much in common and may share the same neural substrate.

Discrimination analysis: decoding of temporally integrated shape in visual cortex

Next, we extended our discrimination analysis to include all of our predefined ROIs. For each ROI, we first applied the same classification methods as described above to calculate the shape-decoding confusion matrix. We then assessed classification performance by calculating the correct classification score (the mean diagonal score in the confusion matrix). We also used the standard correlation technique, computing the correlation between the patterns of activation for each pair of stimuli across viewing conditions (and across different full-image runs). We calculated the excess correlation (for same vs different stimuli) by subtracting the average correlation for different shapes from the correlation level for the same shape (average diagonal score minus average off-diagonal score in the correlation matrix). Figure 5 shows the resulting classification accuracy (top) and the excess correlation (bottom) in the various visual cortical ROIs. Statistically significant performance (i.e., classification performance greater than chance level 11% or excess correlation above zero) is denoted by asterisks (corresponding *t*- and *p*-values are listed in Table 1). Left (“across slits”) and middle (“slit vs full viewing”) panels depict the degree of viewing-invariant shape discrimination in the different ROIs. The right panel provides a quantitative benchmark for this discrimination (when the whole shape is seen at its full in “full-viewing” runs).

The crucial generalization test is the one in which shape discrimination is maintained although the shape is viewed across an orthogonal slit orientation to the one used in the training set.

Highly significant shape discrimination across different slit conditions (Fig. 5, left) was found using both methods in MT+ and higher-level object-selective regions cIPS and LOC, as well as LOC subregions pFs and LO. In addition, significant excess correlation was found in V4, V3ab, and even V3. However, V1, V2, and LO-1 failed in slit-invariant shape discrimination. In contrast, in two other tests in which generalization of shape information may be partially (Fig. 5, middle) or substantially (Fig. 5, right) attributed to shared retinal input, both the classification accuracy and the excess correlation were significant in all regions, including early visual cortical areas (V1, V2, and V3, termed collectively EVC).

To corroborate the existence of viewing-invariant shape representations in MT+, cIPS, and LOC, we applied a two-way ANOVA. The factors were ROI identity (V1–V3, V3ab, MT+, cIPS, V4, LO-1, and LOC) and a three-level factor that reflects the type of generalization (1: across slit orientation, 2: across slit-limited and full viewing, and 3: across different full-viewing runs). These levels are defined by the proportion of shape information shared on the retina in each generalization test. If a particular ROI carries viewing-invariant information about whole shape, then this factor should not influence the ROI’s shape decoding. However, the impact of the factor should be significant if shape information in the ROI is viewing condition dependent. A significant ROI × generalization type interaction showed that areas differ in their degree of generalization across viewing conditions (classification: $F_{(16,240)} = 21.7$, $p = 5 \times 10^{-38}$; excess correlation: $F_{(16,240)} = 26.5$, $p = 3 \times 10^{-44}$). Subsequent simple-main-effects analysis confirmed that LOC, cIPS, and MT+ were the least sensitive/not sensitive to this factor (a one-way ANOVA per ROI: classification: $F_{(2,30)} = 7.6, 7.4, 5.0$; $p = 0.047, 0.053, 0.254$; excess correlation: $F_{(2,30)} = 11.3, 8.9, 4.0$; $p = 0.005, 0.020, 0.486$, respectively, Sidák corrected). In contrast, V1, V2, V3, LO-1, and V4 were highly sensitive (classification: $F_{(2,30)} = 59.8, 43.3, 47.4, 32.8, 23.6$; $p = 8 \times 10^{-10}, 3 \times 10^{-8}, 1 \times 10^{-8}, 6 \times 10^{-7}, 2 \times 10^{-5}$; excess correlation: $F_{(2,30)} = 58.0, 46.3, 38.0, 27.5, 25.8$; $p = 1 \times 10^{-9}, 1 \times 10^{-8}, 1 \times 10^{-7}, 4 \times 10^{-6}, 7 \times 10^{-6}$, respectively).

Importantly, LOC discriminated between the shapes across viewing conditions much better than both MT+ and cIPS (the regions with the second and third highest discrimination capability) (*post hoc* paired *t* tests: “across slits”: classification: $t_{(15)} = 7.3, 8.7$; $p = 2 \times 10^{-5}, 2 \times 10^{-6}$; “slit vs full viewing”: classification: $t_{(15)} = 8.4, 7.9$; $p = 4 \times 10^{-6}, 8 \times 10^{-6}$; excess correlation: $t_{(15)} = 5.5, 6.1$; $p = 5 \times 10^{-4}, 2 \times 10^{-4}$; respectively, Sidák corrected). Within the LOC, the lateral LOC part (LO) was more sensitive to viewing-invariant shape information than the corresponding ventral part (pFs) (“across slits”: classification: $t_{(15)} = 3.4$, $p = 0.015$; excess correlation: $t_{(15)} = 4.9$, $p = 8 \times 10^{-4}$; “slit vs full viewing”: classification: $t_{(15)} = 5.7$, $p = 2 \times 10^{-4}$; excess correlation: $t_{(15)} = 7.0$, $p = 2 \times 10^{-5}$).

To summarize, higher-level visual ROIs in both visual pathways as well as MT+ represent viewing-invariant information about global shape. The strongest shape discrimination capability that is generalized across viewing conditions is seen in LO, the lateral part of LOC. In contrast, viewing-invariant shape information is much more limited in mid-tier dorsal and ventral visual regions and totally absent in EVC. Therefore, although EVC does carry information about the entire shape when the full image is available, this information is a mere copy of the retinal input. Note, that mid-tier visual regions V3ab, V4, and MT+ receive retinotopically organized inputs from EVC (Wandell et al., 2007; Amano et al., 2009). Conversely, they show some sensitivity to slit-invariant (nonretinotopic) information. Therefore, slit-specific

Figure 5. Shape discrimination in visual cortex across viewing conditions. Cross-validated above-chance classification accuracy (top) and excess correlation for same versus different shapes (bottom) in different visual ROIs (averaged across participants). Left inset, Schematic representation of a shape classification confusion matrix. White (diagonal) matrix entries represent correct classifications, whereas black (off-diagonal) matrix entries are misclassifications (see example in Fig. 4). The excess correlation is the difference between the main diagonal and off-diagonal correlation levels (in the correlation matrix). Right inset, Color code and location of the representative participant's ROIs. Left, middle, and right panels, across slit orientation, across slit-limited and full viewing condition, and across different runs of the same full-image condition, respectively. LOC (and its lateral part LO in particular) carries the most significant viewing-invariant shape information. Error bars indicate SEM. Statistically significant shape discrimination is denoted by asterisks. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Other details are as in Figure 4.

and slit-invariant visual representations are likely to coexist in these regions, showing a transition from the slit-dependent analysis of visible stimulus fragments to the inferred global shape representation.

RSA: relating shape perception and brain responses to shapes

Our next goal was to relate the structure of neural shape space (the degree of activation pattern similarity between any two shapes in a specific area) to the structure of perceptual shape space as reported by participants in perceptual similarity judgment test. This could be done by applying RSA (Kriegeskorte et al., 2008).

To do so, we used group mean perceptual distances between the shapes (i.e., dissimilarity measures) estimated in the two behavioral sessions performed before and after scanning (the distances were averaged across the sessions and participants; see Fig. 6A, top, and Materials and Methods). In addition, to assess the degree of dissimilarity between brain activation patterns for each pair of shapes, the off-diagonal scores in the classification-confusion (or correlation) matrices were converted into dissimilarity (or distance) measures (see Materials and Methods). This yielded two

RDMs for each generalization test in each individual ROI (for an example of such an RDM, see Fig. 6A, bottom).

A close match between the perceptual dissimilarity matrix and a particular ROI RDM may indicate that the ROI's shape representations are relevant for the perceptual outcome (although this is obviously only correlative evidence). To test this correspondence, we calculated the proportion of variance (R^2) in the pattern activation RDMs that can be accounted for by the perceived shape dissimilarity. Figure 6B illustrates the resulting estimates in different ROIs for the various generalization tests (for the significance of these estimates, see Table 1).

The metric properties of the shape perceptual space accounted for 18–27% of variance in “across-slits” and “slit vs full viewing” LOC RDMs (Fig. 6B, left and middle, respectively). This proportion was much lower in MT+, cIPS, and mid-tier dorsal and ventral stream regions V3ab, V4, and LO-1 (5–13%). The accounted variance was also significant in EVC when some of the shape was shared on the retina across viewing conditions (i.e., in “slit vs full viewing” test; Fig. 6B, middle), but not when this amount was negligible (i.e., in “across-slits” test; Fig. 6B, left).

Table 1. Statistical analysis of shape discrimination and perceptual–neural correspondence for the standard ROI set

Test	Variable	ROI	Across slits		Slit vs full viewing		Full viewing	
			<i>t</i>	<i>p</i> -corrected	<i>t</i>	<i>p</i> -corrected	<i>t</i>	<i>p</i> -corrected
1	Classification performance	V1	2.1	0.979	5.2	0.007	8.2	4×10^{-5}
		V2	3.3	0.259	6.9	4×10^{-4}	7.33	2×10^{-4}
		V3	2.2	0.935	8.7	2×10^{-5}	7.5	1×10^{-4}
		V3ab	3.8	0.107	6.4	8×10^{-4}	5.1	0.008
		MT+	6.4	8×10^{-4}	4.7	0.018	6.1	0.001
		cIPS	5.9	0.002	6.1	0.001	5.3	0.006
		V4	4.0	0.077	7.4	2×10^{-4}	5.9	0.002
		LO-1	3.3	0.255	4.3	0.041	6.5	6×10^{-4}
		LOC	11.7	4×10^{-7}	8.9	1×10^{-5}	7.8	7×10^{-5}
		LO	8.6	2×10^{-5}	8.3	3×10^{-5}	7.30	2×10^{-4}
		pFs	6.0	0.002	5.6	0.003	5.0	0.010
		V1	0.5	0.999	7.04	3×10^{-4}	7.7	9×10^{-5}
		V2	2.8	0.611	7.8	8×10^{-5}	7.1	2×10^{-4}
2	Excess correlation	V3	4.6	0.021	9.7	5×10^{-6}	6.7	5×10^{-4}
		V3ab	5.6	0.004	6.9	3×10^{-4}	5.86	2×10^{-3}
		MT+	5.9	0.002	6.99	3×10^{-4}	6.1	1×10^{-3}
		cIPS	6.1	0.001	8.0	6×10^{-5}	7.2	2×10^{-4}
		V4	5.0	0.011	9.5	7×10^{-6}	5.95	2×10^{-3}
		LO-1	2.8	0.620	5.1	8×10^{-3}	6.0	2×10^{-3}
		LOC	11.4	6×10^{-7}	10.7	1×10^{-6}	7.6	1×10^{-4}
		LO	10.9	1×10^{-6}	10.9	1×10^{-6}	7.8	8×10^{-5}
		pFs	7.1	2×10^{-4}	7.2	2×10^{-4}	5.8	2×10^{-3}
		V1	2.0	0.982	9.3	9×10^{-6}	8.9	1×10^{-5}
		V2	3.6	0.154	5.4	0.005	5.5	0.004
		V3	2.3	0.903	4.9	0.012	4.6	0.023
		V3ab	5.0	0.010	4.9	0.013	5.6	0.004
3	Explained variance (R^2) in classification-based RDMs	MT+	4.2	0.045	4.5	0.028	3.6	0.155
		cIPS	3.6	0.145	3.8	0.113	4.1	0.066
		V4	3.8	0.106	5.0	0.011	4.0	0.070
		LO-1	4.0	0.074	4.6	0.022	6.7	5×10^{-4}
		LOC	8.4	3×10^{-5}	11.2	8×10^{-7}	6.2	0.0012
		LO	8.1	5×10^{-5}	8.9	1×10^{-5}	6.0	0.002
		pFs	4.3	0.040	3.8	0.113	4.1	0.058
		V1	2.5	0.822	8.5	3×10^{-5}	17.0	2×10^{-9}
		V2	3.6	0.145	7.1	3×10^{-4}	11.1	8×10^{-7}
		V3	4.4	0.031	10.1	3×10^{-6}	8.9	2×10^{-5}
		V3ab	7.0	3×10^{-4}	7.9	6×10^{-5}	8.0	6×10^{-5}
		MT+	8.3	4×10^{-5}	9.0	1×10^{-5}	5.59	0.0034
		cIPS	5.2	0.007	6.9	3×10^{-4}	5.58	0.0035
4	Explained variance (R^2) in correlation-based RDMs	V4	5.7	0.003	7.8	7×10^{-5}	7.1	2×10^{-4}
		LO-1	7.0	3×10^{-4}	6.2	0.0011	7.5	1×10^{-4}
		LOC	18.4	7×10^{-10}	25.4	6×10^{-12}	11.6	4×10^{-7}
		LO	14.5	2×10^{-8}	18.5	6×10^{-10}	11.1	8×10^{-7}
		pFs	4.9	0.012	6.7	5×10^{-4}	4.6	0.024

Statistics were performed based on 16 participants. Shown are shape discrimination across slits, across slit and full-image condition, and across different runs of the same full-image condition. *p*- and *t*-values correspond to one-sample two-tailed *t* test of: (1) classification accuracy against chance level, (2) excess correlation against zero, (3) explained variance (R^2) in participants' classification-based RDMs (accounted for by perceived-shape dissimilarity), and (4) correlation-based RDMs against zero. *p*-values are corrected for multiple comparisons using Sidák's correction.

Next, we applied the same two-way ANOVA as before with the accounted variance as the dependent variable. Similar to the previous results, the interaction term (ROI $\times$ generalization type) was highly significant (classification RDMs: $F_{(16,240)} = 10.2$, $p = 2 \times 10^{-19}$; correlation-RDMs: $F_{(16,240)} = 27.6$, $p = 2 \times 10^{-45}$). Subsequent analysis of simple main effects showed that the three-level factor that reflects the type of generalization was not significant in LOC, cIPS, MT+, and V3ab (a one-way ANOVA per ROI: classification RDMs: $F_{(2,30)} = 0.1, 0.2, 0.1, 2.8$, $p = 0.99, 0.99, 0.99, 0.820$; correlation RDMs: $F_{(2,30)} = 0.5, 0.1, 0.2, 5.3$, $p = 0.99, 0.99, 99, 0.204$, respectively). Therefore, the degree of correspondence between the perceived shape dissimilarity and the activity pattern dissimilarity in these regions does not depend on viewing condition. In contrast, this factor was significant in V1–V3 and V4/LO-1, indicating a dependence on the viewing condition (V1, V2, and V3: classification RDMs: $F_{(2,30)} = 56.7, 21.4, 13.7$; $p = 1 \times 10^{-9}, 4 \times 10^{-5}, 0.0013$; correlation RDMs:

$F_{(2,30)} = 138.4, 72.6, 38.0$, $p = 2 \times 10^{-14}, 7 \times 10^{-11}, 1 \times 10^{-7}$, respectively; V4 and LO-1: correlation RDMs: $F_{(2,30)} = 10.0, 12.0$; $p = 0.011, 0.003$, respectively).

Consistent with the shape discrimination analysis, the estimated variance in viewing-invariant LOC RDMs was much higher than in MT+ and cIPS RDMs (*post hoc* paired *t* test: “across slits”: classification RDMs: $t_{(15)} = 5.74, 5.68$; $p = 3.1 \times 10^{-4}, 3.5 \times 10^{-4}$; correlation RDMs: $t_{(15)} = 6.2, 8.5$; $p = 1 \times 10^{-4}, 3 \times 10^{-6}$; “slit vs full-viewing”: classification RDMs: $t_{(15)} = 5.3, 7.2$; $p = 7 \times 10^{-4}, 3 \times 10^{-5}$; correlation RDMs: $t_{(15)} = 8.6, 12.8$, $p = 3 \times 10^{-6}, 1 \times 10^{-8}$, respectively). Finally, we found that, within LOC, LO was superior to pFs in the percentage variance explained within viewing-invariant RDMs. However, this was evident both when generalizing across different slit-viewing conditions (classification and correlation RDMs: $t_{(15)} = 2.84, 4.11$, $p = 0.049, 0.004$, respectively) and when using the full images (classification and correlation RDMs: $t_{(15)} = 3.56$,

Figure 6. Perceived shape dissimilarity predicts activity pattern dissimilarity across viewing conditions. **A**, Top, Group mean perceived shape dissimilarity. Bottom, Individual example of activation pattern dissimilarity in LO (across slits orientations). Color code: the degree of perceptual (top) or activation pattern (bottom) dissimilarity for each shape pair. **B**, Group mean proportion of variance in participants' activity pattern dissimilarity accounted for by the perceived shape dissimilarity in the different ROIs. The values are the variance explained in either classification-based (top) or correlation-based (bottom) pattern dissimilarity. Their statistical significance is indicated by asterisks. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Error bars indicate SEM. Perceived shape dissimilarity corresponds well to the activation pattern dissimilarity in LOC (and LO in particular) regardless of viewing condition. In contrast, in the low-level visual areas, this correspondence is highest in the full-viewing conditions, but is greatly reduced as the degree of shared retinal input becomes lower.

4.14, $p = 0.011$, 0.003, respectively). We therefore suggest that LO superiority is not specific to temporal shape integration.

To summarize, both dorsal and ventral stream regions show a correspondence between the perceived shape dissimilarity and the viewing-invariant activity pattern dissimilarity. Nevertheless, of all cortical regions tested, the viewing-invariant shape space based on the patterns of activity in LOC best matches the whole-shape perceptual space. This finding is complementary to the decoding results, but goes far beyond the question of shape discriminability: it points to candidate visual regions that are likely to be involved in the integrated shape percept (although this is obviously only circumstantial evidence).

Shape decoding and RSA in fixed-volume ROIs

Because ROI size can affect MVPA results, we repeated the analyses with the nine newly defined ROIs using a uniform fixed size. We applied an activation-based voxel ranking to restrict the size of the ROIs. Some ROIs were combined (due to their small sizes in some participants). Therefore, the new set comprised nine ROIs in each of our 16 participants: V1, V2 + V3, V3ab, V4 + LO-1, MT+, cIPS, LOC, LO, and pFs. We divided the data using two iteration loops. In the "outer" loop, one vertical-slit and one horizontal-slit run was chosen for ROI voxel selection. This was done iteratively using the eight possible choices of the one run (e.g., first vertical and first horizontal run were chosen in the first iteration, etc.). The remaining seven runs were further divided to two subgroups for MVPA using the "inner" loop. For the two runs chosen for voxel selection, a single t -value for each voxel was calculated using the contrast: all slit-viewed shapes versus rest periods (regardless of a slit orientation). Next, we took the 100 most

active voxels that responded best to the shapes. Note that, in each iteration, a relatively small part of the data (1/8) was used for voxel selection, ensuring independence between voxel selection and MVPA. Therefore, the chosen 100 voxels in each iteration were somewhat different for each fixed-size ROI. The mean degree of overlap between the voxels selected at the different iterations was 59%. The MVPA results were averaged across all eight iterations.

The results of discrimination and RSA analyses are shown in Figure 7, **A** and **B**, respectively (for significance of these results, see Table 2). As Figure 7 shows, the new results largely confirm the results of the previous ROI analysis and, importantly, highlight the superiority of LOC in viewing-invariant shape representation upon the other ROIs. This is reflected in both slit-invariant shape discrimination and the degree of correspondence between the perceived shape dissimilarities and slit-invariant pattern dissimilarities (comparison with the regions with the second and third highest discrimination capability, MT+ and cIPS: paired t test: classification: $t_{(15)} = 7.6, 7.1$; $p = 6 \times 10^{-6}, 1 \times 10^{-5}$; excess correlation: $t_{(15)} = 5.7, 5.2$; $p = 2 \times 10^{-4}, 4 \times 10^{-4}$; classification RDMs: $t_{(15)} = 5.0, 8.3$; $p = 6 \times 10^{-4}, 2 \times 10^{-6}$; correlation RDMs: $t_{(15)} = 5.9, 8.8$; $p = 1 \times 10^{-4}, 1 \times 10^{-6}$, respectively).

Lack of shape-decoding information during randomized temporal order presentation

Of concern was that the slit-invariant shape information present in LOC might be due to some feature in the image (e.g., specific curvature) that was retained across the radical slit-angle change. The images were symmetric and receptive fields increase dramatically along the hierarchy from V1 to LOC, possibly allowing for such an explanation of our slit-invariant decoding capabilities,

Figure 7. Shape discrimination and RSA in fixed-volume ROIs. **A**, Cross-validated classification accuracy (top) and excess correlation (bottom) for same versus different shapes across participants in each ROI. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$. Other details are as in Figures 5 and 6. **B**, Group mean proportion of variance in the activation pattern dissimilarity that can be accounted for by perceived shape dissimilarity in different fixed-volume ROIs for different generalization tests. To better visualize the differences between the ROIs, the proportion of variance is plotted as a function of shape information that is shared on the retina across viewing conditions in the different generalization tests (i.e., 1, 10, and 100% on log scale). The results with the fixed-size ROIs corroborate the previous analysis shown in Figures 5 and 6.

especially in higher-order areas. To test this explicitly, we ran an additional control experiment (in 11 of 16 participants), presenting the same set of slit-dependent shape views in a shuffled temporal order. This “random temporal order” condition never gave rise to the percept of a whole shape, in contrast to the stimuli presented in their original temporal order (Morgan et al., 1982). Although shape features can potentially still be extracted (from single shape fragments), they are perceived as separate events without regard to the global shape information that potentially exists in their correct temporal sequence.

We hypothesized that, if a particular cortical region truly represents a whole, temporally integrated shape, then it would fail to show across-slit generalization in such conditions. Figure 8 shows the results of discrimination (Fig. 8, top) and RSA analyses (bottom) for the standard ROI set (Fig. 8A) and fixed-volume ROIs (Fig. 8B). A two-way ANOVA was run to examine the effect of sliver presentation order and ROI identity on the estimates of shape information. A significant temporal order $\times$ ROI identity interaction was found when testing both the standard-set ROIs and fixed-volume ROIs (classification: $F_{(8,80)} = 12.9$, $p = 9 \times 10^{-12}$; $F_{(6,60)} = 13.6$, $p = 1 \times 10^{-9}$; excess correlation: $F_{(8,80)} = 9.8$, $p = 2 \times 10^{-9}$; $F_{(6,60)} = 18.1$, $p = 8 \times 10^{-12}$; classification RDMs: $F_{(8,80)} = 5.7$, $p = 9 \times 10^{-6}$; $F_{(6,60)} = 9.7$, $p = 2 \times 10^{-7}$; correlation RDMs: $F_{(8,80)} = 14.4$, $p = 9 \times 10^{-13}$; $F_{(6,60)} = 12.4$, $p = 5 \times 10^{-9}$, respectively). *Post hoc* paired *t* tests revealed that the amount of shape information in LOC and LO ROIs (in the standard set) dropped significantly when the successive shape views were presented in a random temporal order (Fig. 8A, left)

compared with their correct temporal order (right). Importantly, this was evident for both shape discrimination (LOC and LO: classification: $t_{(10)} = 8.6$, 6.0 ; $p = 2 \times 10^{-5}$, 5×10^{-4} ; excess correlation: $t_{(10)} = 11.5$, 8.8 ; $p = 2 \times 10^{-6}$, 2×10^{-5} , respectively) and the degree of correspondence between the perceived shape dissimilarities and slit-invariant pattern dissimilarities (LOC and LO: classification RDMs: $t_{(10)} = 5.2$, 6.7 ; $p = 0.0016$, 2×10^{-4} ; correlation RDMs: $t_{(10)} = 7.2$, 8.3 ; $p = 1 \times 10^{-4}$, 3×10^{-5} , respectively). Furthermore, the effect was the same for the fixed-size ROI set (Fig. 8B; LOC and LO: classification: $t_{(10)} = 8.0$, 5.2 ; $p = 1 \times 10^{-4}$, 0.005 ; excess correlation: $t_{(10)} = 7.7$, 7.2 ; $p = 2 \times 10^{-4}$, 3×10^{-4} ; classification RDMs: $t_{(10)} = 11.3$, 7.1 ; $p = 6 \times 10^{-6}$, 4×10^{-4} ; correlation RDMs: $t_{(10)} = 8.1$, 8.4 ; $p = 1 \times 10^{-4}$, 9×10^{-5} , respectively).

The effect is likely related to the fact that a global shape percept cannot be formed in the “random order” condition (despite the fact that the same shape features were presented and attended). Temporal integration is therefore probably an essential element for global shape perception in our slit-viewing conditions. We conclude that across-slit generalization of shape information seen in LOC and LO in the original presentation condition does not reflect mere generalization of low-level feature-related information (available within a slit).

Recovered shape information in higher-level ventral stream regions within and beyond LOC

LOC is a vast cortical expanse including a number of distinguishable ventral subregions. We examined the distribution of recov-

Table 2. Statistical analysis of shape discrimination and perceptual-neural correspondence for the fixed-volume ROIs

Test	Variable	ROI(s)	Across slits		Slit vs full viewing		Full viewing	
			<i>t</i>	<i>p</i> -corrected	<i>t</i>	<i>p</i> -corrected	<i>t</i>	<i>p</i> -corrected
1	Classification performance	V1	1.0	0.999	7.7	7×10^{-5}	8.2	4×10^{-5}
		V2 and V3	1.9	0.991	9.3	7×10^{-6}	7.4	1×10^{-4}
		V3ab	5.4	0.004	6.3	7×10^{-4}	5.2	0.006
		MT+	7.3	1×10^{-4}	5.4	0.004	7.1	2×10^{-4}
		dIPS	5.7	0.002	7.66	8×10^{-5}	5.6	0.003
		V4 and LO-1	6.9	3×10^{-4}	7.68	8×10^{-5}	6.9	3×10^{-4}
		LOC	13.0	8×10^{-8}	9.4	6×10^{-6}	7.9	5×10^{-5}
		LO	9.2	8×10^{-6}	8.9	1×10^{-5}	7.3	1×10^{-4}
		pFs	7.3	1×10^{-4}	5.9	0.002	5.8	0.002
2	Excess correlation	V1	−0.5	0.999	7.8	7×10^{-5}	8.3	3×10^{-5}
		V2 and V3	3.8	0.085	8.5	2×10^{-5}	7.0	2×10^{-4}
		V3ab	5.30	0.005	7.2	2×10^{-4}	5.6	0.003
		MT+	5.9	0.0015	6.2	9×10^{-4}	6.1	0.0011
		dIPS	6.5	6×10^{-4}	8.3	3×10^{-5}	7.3	1×10^{-4}
		V4 and LO-1	5.26	0.005	7.1	2×10^{-4}	6.4	7×10^{-4}
		LOC	10.1	2×10^{-6}	10.3	2×10^{-6}	8.0	5×10^{-5}
		LO	9.9	3×10^{-6}	10.9	8×10^{-7}	7.6	9×10^{-5}
		pFs	7.2	2×10^{-4}	6.8	4×10^{-4}	5.7	0.002
3	Explained variance (R^2) in classification-based RDMs	V1	2.8	0.505	8.2	4×10^{-5}	10.7	1×10^{-6}
		V2 and V3	4.3	0.035	7.7	7×10^{-5}	6.9	3×10^{-4}
		V3ab	5.9	0.0015	6.6	4×10^{-4}	4.8	0.013
		MT+	4.8	0.013	5.7	0.002	3.7	0.104
		dIPS	4.0	0.066	4.4	0.027	4.0	0.067
		V4 and LO-1	3.0	0.379	5.9	0.0014	5.1	0.007
		LOC	16.6	2×10^{-9}	16.8	2×10^{-9}	6.8	3×10^{-4}
		LO	13.2	7×10^{-8}	11.1	7×10^{-7}	7.0	2×10^{-4}
		pFs	4.0	0.060	5.3	0.004	5.0	0.008
4	Explained variance (R^2) in correlation-based RDMs	V1	3.0	0.362	9.2	8×10^{-6}	19.5	3×10^{-10}
		V2 and V3	3.4	0.191	9.9	3×10^{-6}	11.0	8×10^{-7}
		V3ab	6.4	6×10^{-4}	10.2	2×10^{-6}	7.1	2×10^{-4}
		MT+	8.6	2×10^{-5}	9.2	8×10^{-6}	5.76	0.002
		dIPS	4.7	0.016	7.0	2×10^{-4}	5.78	0.002
		V4 and LO-1	6.7	4×10^{-4}	11.2	6×10^{-7}	8.7	2×10^{-5}
		LOC	18.5	5×10^{-10}	21.1	8×10^{-11}	10.6	1×10^{-6}
		LO	14.8	1×10^{-8}	18.0	8×10^{-10}	11.4	5×10^{-7}
		pFs	4.9	0.011	7.3	1×10^{-4}	4.7	0.017

Statistics were performed based on 16 participants. Other details are as in Table 1.

ered shape information in higher-level visual regions within and beyond LOC. For that purpose, we ran an additional face/body localizer (in 11 of our original 16 participants) and defined a new set of ROIs (i.e., EBA, FBA, FFA, and OFA). Face/body ROIs were restricted to voxels that do not overlap with LOC. Furthermore, we divided each individual LOC ROI into two parts. All overlapping voxels (between the LOC and each category-specific ROI) were grouped and labeled “shared LOC.” The rest of the LOC voxels were defined as “nonshared LOC.” We then extended our analyses to these ROIs in an effort to localize where is the information about integrated shape present within the higher-level regions of the ventral visual pathway. Note that our study focuses on patterns of activation related exclusively to object shape and therefore applies to artificial, unfamiliar objects (to dissociate these patterns from categorical, semantic representations, as in other fMRI studies; Op de Beeck et al., 2008; Vernon et al., 2016). We therefore expected that such shapes might be encoded in nonspecialized, higher-level ventral stream regions (i.e., “nonshared LOC”).

Figure 9A represents the results of shape decoding (i.e., classification accuracy) and RSA (i.e., the estimated variance in ROIs’ RDMs, the variance was averaged across classification- and correlation-based RSA). Both the group mean decoding accuracy and accounted variance were significantly higher in “nonshared LOC” compared with a body/face ROI (with the highest discrim-

ination capability/accounted variance). This was the case in each generalization test (paired *t* test: “across slits”: classification: $t_{(10)} = 7.2, p = 8 \times 10^{-5}$; RDMs: $t_{(10)} = 2.9, p = 0.045$; “slit vs full viewing”: classification: $t_{(10)} = 5.9, p = 5 \times 10^{-4}$; RDMs: $t_{(10)} = 6.4, p = 2 \times 10^{-4}$; “full viewing”: classification: $t_{(10)} = 5.3, p = 0.00103$; RDMs: $t_{(10)} = 4.0, p = 0.007$). To ensure that the result does not depend on the ROIs’ size, we pooled all body/face voxels into one ROI and limited the amount of voxels in “shared LOC,” “nonshared LOC,” and combined body/face ROI to 100 voxels (using the same procedure described previously; see Materials and Methods). As Figure 9B shows, “nonshared LOC” is still clearly better than the categorical ROI in its shape information regardless of generalization test (“across slits”: classification: $t_{(10)} = 3.9, p = 0.008$; RDMs: $t_{(10)} = 5.0, p = 0.0015$; “slit vs full viewing”: classification: $t_{(10)} = 4.3, p = 0.005$; RDMs: $t_{(10)} = 4.8, p = 0.002$; “full viewing”: classification: $t_{(10)} = 5.9, p = 4 \times 10^{-4}$; RDMs: $t_{(10)} = 4.6, p = 0.003$).

Therefore, both conventional and temporally integrated shape representations are embedded in the activation profiles of the least specialized (i.e., noncategorical) parts of LOC.

Whole-brain MVPA

It is possible that there are other cortical regions that carry information regarding a temporally integrated shape beyond the classical ROIs tested. We applied a volume-based searchlight MVPA

Figure 8. Lack of shape decoding information during randomized temporal order presentation. Top, Shape discrimination (classification accuracy) across different slit-viewing conditions during the shuffled presentation (“random”) and the original order (“correct”) of presentation. The data are shown for the standard ROIs (**A**) and the fixed-volume ROIs (**B**). Note the dramatic drop in decoding information between the “random” and “correct” order, especially in LO and LOC. The same general pattern can be seen in the analysis of excess correlation for the same shape when compared with different shapes. Bottom, Proportion of variance in the activation pattern dissimilarity that can be accounted for by perceived shape dissimilarity using classification-based RDMs and correlation-based RDMs. The same drop in the amount of information regarding the various shapes is evident in the random temporal order condition. Therefore, low-level (feature-based) cues per se are not sufficient to generate the slit-invariant shape information evident in LOC. Statistical significance of the shape information estimates is indicated by asterisks. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

that makes no assumption about the location of the shape-specific information (Kriegeskorte et al., 2006). We searched iteratively through a brain with a cubic search window comprising 125 voxels (i.e., a total volume of $15\text{ mm} \times 15\text{ mm} \times 15\text{ mm} = 3375\text{ mm}^3$) and then performed SVM classification analysis as in the ROI analysis. A classifier was trained to distinguish between

shapes under one slit orientation and its generalization performance was tested under the other orientation.

Figure 10 depicts the groupwise statistical maps showing above-chance classification accuracy (Fig. 10A) and significant variance in across-slit activation pattern dissimilarity accounted for by the perceived shape dissimilarity (Fig. 10B). The whole-

Figure 9. Recovered shape information in higher-level ventral stream regions within and beyond LOC. **A, B**, Average shape-decoding levels (classification accuracy) and the degree of matching between the perceptual and pattern activation dissimilarity (percentage variance explained) in the various ROIs that comprise higher-level ventral stream areas. The color code indicates different ROIs. Voxels that were mutual to LOC and specialized ROIs are denoted in gray ("shared LOC"). The data in **A** are from the standard ROI set and those in **B** are from the fixed-size ROI set. The largest amount of slit-invariant shape information is seen in the least specialized higher-level ventral stream regions (i.e., "nonshared LOC"). Statistical significance of the estimates is indicated by asterisks. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Figure 10. Slit-invariant shape representation in the whole brain. **A, B**, Group searchlight statistical t maps are shown on the cortical surface of one participant from a posterior ventral and dorsal (in inset) views [searchlight size: $5 \times 5 \times 5$ voxels; $q(\text{FDR}) < 0.001$]. The classifier was trained to distinguish between shapes in one slit condition and its generalization performance was tested in the orthogonal slit condition. **A**, Map showing the regions in which classification accuracy was significantly greater than chance level across participants. **B**, Map of the regions in which there was a significant correspondence between perceived shape judgments and the patterns of activation evoked by these shapes during slit viewing. Formally, these were areas in which the proportion of variance in activation pattern dissimilarity matrix accounted for by the perceived shape dissimilarity was significantly greater than zero across participants. Color code indicates the degree of statistical significance. **C**, Group ROI probability map for LOC ROI in which all individual LOC ROIs were superimposed yielding a probability map. Warmer colors represent standard voxels that were included in a greater proportion of the individual ROIs. Solid black line in **A** and **B** represents isoproportionality contours extracted from the map. The regions within the contour boundaries enclose standard voxels that are part of at least two individual ROIs (i.e., the inclusion probability is $\geq 13\%$). As can be seen, both the groupwise classification accuracy (**A**) and accounted variance (**B**) are most significant within the contour boundaries, thus corroborating strong LOC sensitivity to the temporally integrated shape information.

brain analysis does not reveal any additional loci carrying substantial information about the recovered shape: Both groupwise classification accuracy and accounted variance are most significant in similar regions of the lateral occipitotemporal cortex (x - y - z Talairach mean coordinates: left: -40 , -71 , -1 and -39 , -74 , -4 , respectively; right: 39 , -69 , -3 and 39 , -72 , -5 , respectively). Importantly, the coordinates of these regions in standard space closely correspond to either the group mean LO coordinates (left and right hemisphere: -42 ± 3 , -71 ± 5 , -5 ± 6 and 41 ± 3 , -70 ± 5 , -5 ± 5 , respectively) or LOC coordinates (-41 ± 2 , -66 ± 4 , -7 ± 5 and 40 ± 2 , -65 ± 4 , -7 ± 3 , respectively). A probabilistic map of LOC voxels across participants (Fig. 10C) allows visualization of the voxels' location in the brain. The overlap between the LOC voxels and the searchlight clusters is shown here by depicting the probabilistic map boundaries superimposed on the searchlight statistical maps (solid black contour in Fig. 10A,B). The spatial correspondence between the searchlight clusters and LOC corroborates the results of our ROI analysis about the crucial role of this region in representation of slit-invariant global shape information.

Discussion

We demonstrate here that ventral stream object representations maintain their selectivity for object shape when shape structure can only be inferred by integration over time (e.g., in slit-viewing conditions). The strength of the present study is that it can distinguish between temporally integrated shape information and "trivial" shape information available already in the retinal input. This was achieved by manipulating slit orientation and testing for generalization of shape representation across slit conditions. Our results show compellingly that LOC best meets this requirement (Figs. 5, 7, 10). Importantly, slit-invariant shape information in LOC was correlated with the global shape perception (Figs. 6, 7, 10). Shape information in LOC was almost totally gone when the shape percept was eliminated by shuffling the order of slit-dependent shape views (Fig. 8). Together, these results suggest that LOC represents the outcome of the shape temporal integration and is likely to mediate the percept of the integrated shape. Finally, the slit-invariant, shape-specific patterns of activation also matched those elicited by conventional whole-shape images (which were assessed independently, in full-image runs after slit runs). Importantly, the two representation forms (of full shape images vs temporally integrated ones) emerge from largely different integrative processes. However, our results provide evidence that they converge on a common representation in LOC (Figs. 5, 6, 7). Therefore, LOC holds an abstract representation of global shape invariant, not only to scale, position, viewpoint, etc. (Grill-Spector et al., 1998, 1999; Kourtzi and Kanwisher, 2001; Vuilleumier et al., 2002; Weigelt et al., 2007; Eger et al., 2008; Vinberg and Grill-Spector, 2008; Cichy et al., 2011), but also to the type of integration by which it was obtained.

LOC has been found to mediate object spatial completion in stationary (partially occluded) visual scenes (Lerner et al., 2002; Murray et al., 2004; Tang et al., 2014). Unlike the spatial completion process, temporal shape integration requires recovery of the global motion vector for shape reconstruction (Shimojo and Richards, 1986) and thus may engage mechanisms located separately from the conventional shape representations in LOC. However, there were some indications that LOC might also be involved in temporal integration of object shapes. A previous study showed that LOC was highly active when slit-viewed drawings of familiar objects were perceived as an integrated whole and this activity decreased when distortion of the objects' contours precluded

recognition (Yin et al., 2002). A recent fMRI study addressed contour integration under slit-viewing conditions. The stimuli were composed of multiple collinear Gabor elements embedded among other randomly oriented Gabor elements (Kuai et al., 2017). The pattern was nonetheless perceived as a straight line (due to the collinearity of Gabor elements) tilted to the left or to the right (according to the spatial arrangement of the aligned Gabor patches). This emergent percept was also present if the stimulus was translating behind a slit. Under slit-viewing conditions, the patterns of activation in LOC allowed significant discrimination between co-linear and randomly oriented Gabors. However, unlike our case, LOC did not show contour specificity: it failed to discriminate between left- and right-tilted contours. Crucially, this failure was not only under slit-viewing conditions, but was also present when using static Gabor pattern displays. This is probably due to the fact that stimuli composed of hundreds of Gabor patches are first and foremost textures rather than objects. Objects are clearly perceived when the integrated contours form a closed shape (e.g., a circle). LOC is defined by its preference for objects compared with textures (Malach et al., 1995). Therefore, it is expected that this region would be "blind" to the internal structure of the textures unless the components form a clear shape.

Our experiment was designed to account for global shape information focusing on the final product of the temporal integration process. We therefore cannot answer the question of how exactly integration was achieved. However, below we describe at what stage of the visual hierarchy the relevant computations for temporal integration may take place and distinguish this process from the extraction of shape in full-viewing conditions:

When a fully viewed object is presented, a mere feedforward (initial) sweep of visual processing is probably sufficient to integrate the local shape elements over space, thus generating a whole-shape percept. This is possible because the complete spatial layout of the object parts is available simultaneously and in the same retinotopic reference frame (DiCarlo and Cox, 2007). Classical models of object perception start with this retinal image as the necessary first step and apply spatial integration of the retinal information at larger and larger scales along the visual hierarchy, culminating in the representation of object identity (Riesenhuber and Poggio, 1999; Serre et al., 2007; Hong et al., 2016; see DiCarlo and Cox, 2007 for a review).

However, in our experimental conditions, the slit-viewed shapes stimulated the same narrow retinal strip over time. This is a serious challenge for those models: a solely space-based mechanism would incorrectly blend the shape fragments presented in succession. Therefore, shape integration must be guided by a process that can potentially integrate information across both space and time. Previous behavioral studies show that coherent shape perception under slit viewing strongly depends on the correct recovery of global shape velocity (Morgan et al., 1982; Casco and Morgan, 1984; Shimojo and Richards, 1986; Sohmiya and Sohmiya, 1994; Rieger et al., 2007; Aydin et al., 2008; Palmer and Kellman, 2014). Obviously, correct estimation of the velocity is critical for predicting the future spatial positions of the occluded shape-parts (Shimojo and Richards, 1986). If spatial positions of these internally maintained shape fragments are updated over time, then the fragments can be integrated with the visible portion of the shape (using Gestalt principles such as good continuation; Palmer et al., 2006). To perform this update successfully, the slit-dependent shape views must be "freed" from retinotopy and encoded in nonretinotopic sensory memory (Ögmen and Herzog, 2016). Importantly, when the spatiotemporal continuity of the stimulus is violated (i.e., in our "random temporal order"

condition), this nonretinotopic reconstruction cannot be established and the percept of both shape and motion are gone.

Our study shows that LOC is likely to mediate this reconstructive behavior. This may not be a great surprise. If shape integration over time depends on the correct assessment of the global shape motion, then it cannot start at the earliest stages of visual hierarchy before the point at which the aperture problem is resolved. Indeed, slit-viewed objects activate EVC regardless of whether they are perceived as a whole or as separate line segments (Yin et al., 2002). Obviously, the patterns of activity in EVC allow discrimination between the various shapes when the extended shape retinal image is available, but they do not carry any noticeable slit-invariant shape information (Figs. 5, 7). Feedback loops to EVC may be used to achieve space-based object completion (Murray et al., 2004; Ban et al., 2013; Muckli et al., 2015). Indeed, neurons in V1 are selective to subjective contours (Peterhans and von der Heydt, 1989), but reconstruction of shape based on temporal integration is probably beyond V1 because it requires shifting the contour position from its retinal location to a new, inferred location using the global velocity estimate.

Human fMRI and transcranial magnetic stimulation studies show that both MT+ and V3a compute global motion and track the location of a moving object even after it becomes hidden behind an occluder (Maus et al., 2010, 2013; Vetter et al., 2015; Chen et al., 2016). These dorsal stream regions are interconnected with various ventral stream regions through the vertical occipital fasciculus (VOF), a major fiber bundle connecting dorsal and ventral parts of human occipital cortex, and adjacent parietal and temporal cortex (Yeatman et al., 2014; Takemura et al., 2016). VOF is therefore likely to provide the communication of signals between ventral stream regions involved in form perception and dorsal stream regions involved in analysis of visual motion (Yeatman et al., 2014). Therefore, MT+ and V3a can potentially mediate the motion-derived information (necessary for object reconstruction under slit-viewing conditions) to LOC. Previous studies showed that LOC is activated by objects solely defined by their common motion (Grill-Spector et al., 1998; Vinberg and Grill-Spector, 2008) and its activity continues even when the motion is stopped as long as the percept of the object persists (Ferber et al., 2003). Possibly, such persistent representations might also allow the integration of visible and occluded shape parts (Palmer et al., 2006) in LOC to represent hidden shape parts in their inferred locations.

Our study also found some degree of slit-invariant shape sensitivity in midlevel regions of both dorsal and ventral pathways (e.g., MT+, V3ab, V4, and LO-1; Figs. 5, 7). Whether temporal integration process is initiated in these regions and finalized in LOC or propagated back to these areas later is unknown. Techniques with finer temporal resolution (e.g., EEG, MEG) are required to understand the complex, hierarchical neural networks underlying this sophisticated process.

References

- Amano K, Wandell BA, Dumoulin SO (2009) Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *J Neurophysiol* 102:2704–2718. [CrossRef Medline](#)
- Anstis SM, Atkinson J (1967) Distortions in moving figures viewed through a stationary slit. *Am J Psychol* 80:572–585. [CrossRef Medline](#)
- Aydin M, Herzog MH, Öğmen H (2008) Perceived speed differences explain apparent compression in slit viewing. *Vision Res* 48:1603–1612. [CrossRef Medline](#)
- Ban H, Yamamoto H, Hanakawa T, Urayama S, Aso T, Fukuyama H, Ejima Y (2013) Topographic representation of an occluded object and the effects of spatiotemporal context in human early visual areas. *J Neurosci* 33:16992–17007. [CrossRef Medline](#)
- Carmines EG, Zeller RA (1979) Reliability and validity assessment. Beverly Hills, CA: California Sage Publications.
- Casco C, Morgan M (1984) The relationship between space and time in the perception of stimuli moving behind a slit. *Perception* 13:429–441. [CrossRef Medline](#)
- Chan AW, Kravitz DJ, Truong S, Arizpe J, Baker CI (2010) Cortical representations of bodies and faces are strongest in commonly experienced configurations. *Nat Neurosci* 13:417–418. [CrossRef Medline](#)
- Chang C, Lin C (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Sys* 2:21–27.
- Chen N, Cai P, Zhou T, Thompson B, Fang F (2016) Perceptual learning modifies the functional specializations of visual cortical areas. *Proc Natl Acad Sci U S A* 113:5724–5729. [CrossRef Medline](#)
- Cichy RM, Chen Y, Haynes JD (2011) Encoding the identity and location of objects in human LOC. *Neuroimage* 54:2297–2307. [CrossRef Medline](#)
- DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. *Trends Cogn Sci* 11:333–341. [CrossRef Medline](#)
- Di Russo F, Pitzalis S, Spinelli D (2003) Fixation stability and saccadic latency in elite shooters. *Vision Res* 43:1837–1845. [CrossRef Medline](#)
- Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. *Science* 293:2470–2473. [CrossRef Medline](#)
- Downing PE, Wiggett AJ, Peelen MV (2007) Functional magnetic resonance imaging investigation of overlapping lateral occipitotemporal activations using multi-voxel pattern analysis. *J Neurosci* 27:226–233. [CrossRef Medline](#)
- Eger E, Ashburner J, Haynes JD, Dolan RJ, Rees G (2008) fMRI activity patterns in human LOC carry information about object exemplars within category. *J Cogn Neurosci* 20:356–370. [CrossRef Medline](#)
- Ejaz N, Hamada M, Diedrichsen J (2015) Hand use predicts the structure of representations in sensorimotor cortex. *Nat Neurosci* 18:1034–1040. [CrossRef Medline](#)
- Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex* 7:181–192. [CrossRef Medline](#)
- Ferber S, Humphrey GK, Vilis T (2003) The lateral occipital complex subserves the perceptual persistence of motion-defined groupings. *Cereb Cortex* 13:716–721. [CrossRef Medline](#)
- Gelis J (2003) A generic geometric transformation that unifies a wide range of natural and abstract shapes. *Am J Bot* 90:333–338. [CrossRef Medline](#)
- Golarai G, Ghahremani DG, Whitfield-Gabrieli S, Reiss A, Eberhardt JL, Gabrieli JD, Grill-Spector K (2007) Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nat Neurosci* 10:512–522. [Medline](#)
- Grill-Spector K, Kushnir T, Edelman S, Itzhak Y, Malach R (1998) Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron* 21:191–202. [CrossRef Medline](#)
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187–203. [CrossRef Medline](#)
- Grill-Spector K, Kushnir T, Hendler T, Malach R (2000) The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci* 3:837–843. [CrossRef Medline](#)
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430. [CrossRef Medline](#)
- Haxby JV, Connolly AC, Guntupalli JS (2014) Decoding neural representational spaces using multivariate pattern analysis. *Annu Rev Neurosci* 37:435–456. [CrossRef Medline](#)
- Hong H, Yamins DL, Majaj NJ, DiCarlo JJ (2016) Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat Neurosci* 19:613–622. [CrossRef Medline](#)
- Huk AC, Dougherty RF, Heeger DJ (2002) Retinotopy and functional subdivision of human areas MT and MST. *J Neurosci* 22:7195–7205. [Medline](#)
- Julian JB, Fedorenko E, Webster J, Kanwisher N (2012) An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* 60:2357–2364. [CrossRef Medline](#)
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311. [Medline](#)
- Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by

- the human lateral occipital complex. *Science* 293:1506–1509. [CrossRef Medline](#)
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Kuai SG, Li W, Yu C, Kourtzi Z (2017) Contour Integration over Time: Psychophysical and fMRI Evidence. *Cereb Cortex* 27:3042–3051. [CrossRef Medline](#)
- Lerner Y, Hendler T, Malach R (2002) Object-completion effects in the human lateral occipital complex. *Cereb Cortex* 12:163–177. [CrossRef Medline](#)
- Liu H, Agam Y, Madsen JR, Kreiman G (2009) Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62:281–290. [CrossRef Medline](#)
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92:8135–8139. [CrossRef Medline](#)
- Maus GW, Weigelt S, Nijhawan R, Muckli L (2010) Does area V3A predict positions of moving objects? *Front Psychol* 1:186. [CrossRef Medline](#)
- Maus GW, Fischer J, Whitney D (2013) Motion-dependent representation of space in area MT+. *Neuron* 78:554–562. [CrossRef Medline](#)
- Misaki M, Kim Y, Bandettini PA, Kriegeskorte N (2010) Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* 53:103–118. [CrossRef Medline](#)
- Morgan MJ, Findlay JM, Watt RJ (1982) Aperture viewing: a review and a synthesis. *Q J Exp Psychol A* 34:211–233. [CrossRef Medline](#)
- Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, Yacoub E (2015) Contextual feedback to superficial layers of v1. *Curr Biol* 25:2690–2695. [CrossRef Medline](#)
- Murray MM, Foxe DM, Javitt DC, Foxe JJ (2004) Setting boundaries: brain dynamics of modal and amodal illusory shape completion in humans. *J Neurosci* 24:6898–6903. [CrossRef Medline](#)
- Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N (2014) A toolbox for representational similarity analysis. *PLoS Comput Biol* 10:e1003553. [CrossRef Medline](#)
- Öğmen H, Herzog MH (2016) A new conceptualization of human visual sensory-memory. *Front Psychol* 7:830. [CrossRef Medline](#)
- Op de Beeck HP, Torfs K, Wagemans J (2008) Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *J Neurosci* 28:10111–10123. [CrossRef Medline](#)
- O'Reilly RC, Wyatte D, Herd S, Mingus B, Jilk DJ (2013) Recurrent processing during object recognition. *Front Psychol* 4:124. [CrossRef Medline](#)
- Palmer EM, Kellman PJ (2014) The Aperture Capture Illusion: misperceived forms in dynamic occlusion displays. *J Exp Psychol Hum Percept Perform* 40:502–524. [CrossRef Medline](#)
- Palmer EM, Kellman PJ, Shipley TF (2006) A theory of dynamic occluded and illusory object perception. *J Exp Psychol Gen* 135:513–541. [CrossRef Medline](#)
- Parks TE (1965) Post-retinal visual storage. *Am J Psychol* 78:145–147. [CrossRef Medline](#)
- Peterhans E, von der Heydt R (1989) Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *J Neurosci* 9:1749–1763. [Medline](#)
- Reddy L, Tsuchiya N, Serre T (2010) Reading the mind's eye: decoding category information during mental imagery. *Neuroimage* 50:818–825. [CrossRef Medline](#)
- Reichert C, Fendrich R, Bernarding J, Tempelmann C, Hinrichs H, Rieger JW (2014) Online tracking of the contents of conscious perception using real-time fMRI. *Front Neurosci* 8:116. [CrossRef Medline](#)
- Rieger JW, Gruschow M, Heinze HJ, Fendrich R (2007) The appearance of figures seen through a narrow aperture under free viewing conditions: effects of spontaneous eye motions. *J Vis* 7:10. [Medline](#)
- Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2:1019–1025. [CrossRef Medline](#)
- Rock I (1981) Anorthoscopic perception. *Sci Am* 244:145–153. [CrossRef Medline](#)
- Schreiber K, Krelberg B (2013) The statistical analysis of multi-voxel patterns in functional imaging. *PLoS One* 8:e69328. [CrossRef Medline](#)
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268:889–893. [CrossRef Medline](#)
- Serre T, Kreiman G, Kouh M, Cadieu C, Knoblich U, Poggio T (2007) A quantitative theory of immediate visual recognition. *Prog Brain Res* 165:33–56. [CrossRef Medline](#)
- Shimojo S, Richards W (1986) “Seeing” shapes that are almost totally occluded: a new look at Parks’s camel. *Percept Psychophys* 39:418–426. [CrossRef Medline](#)
- Sohmiya T, Sohmiya K (1994) What is a crucial determinant in anorthoscopic perception? *Percept Motor Skills* 78:987–998. [Medline](#)
- Takemura H, Rokem A, Winawer J, Yeatman JD, Wandell BA, Pestilli F (2016) A major human white matter pathway between dorsal and ventral visual cortex. *Cereb Cortex* 26:2205–2214. [CrossRef Medline](#)
- Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Stuttgart: Thieme.
- Tang H, Buia C, Madhavan R, Crone NE, Madsen JR, Anderson WS, Kreiman G (2014) Spatiotemporal dynamics underlying object completion in human ventral visual cortex. *Neuron* 83:736–748. [CrossRef Medline](#)
- Vernon RJ, Gouws AD, Lawrence SJ, Wade AR, Morland AB (2016) Multivariate patterns in the human object-processing pathway reveal a shift from retinotopic to shape curvature representations in lateral occipital areas, LO-1 and LO-2. *J Neurosci* 36:5763–5774. [CrossRef Medline](#)
- Vetter P, Grosbras MH, Muckli L (2015) TMS over V5 disrupts motion prediction. *Cereb Cortex* 25:1052–1059. [CrossRef Medline](#)
- Vinberg J, Grill-Spector K (2008) Representation of shapes, edges, and surfaces across multiple cues in the human visual cortex. *J Neurophysiol* 99:1380–1393. [CrossRef Medline](#)
- Vuilleumier P, Henson RN, Driver J, Dolan RJ (2002) Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat Neurosci* 5:491–499. [CrossRef Medline](#)
- Walther A, Nili H, Ejaz N, Alink A, Kriegeskorte N, Diedrichsen J (2016) Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* 137:188–200. [CrossRef Medline](#)
- Wandell BA, Dumoulin SO, Brewer AA (2007) Visual field maps in human cortex. *Neuron* 56:366–383. [CrossRef Medline](#)
- Weigelt S, Kourtzi Z, Kohler A, Singer W, Muckli L (2007) The cortical representation of objects rotating in depth. *J Neurosci* 27:3864–3874. [CrossRef Medline](#)
- Yeatman JD, Weiner KS, Pestilli F, Rokem A, Mezer A, Wandell BA (2014) The vertical occipital fasciculus: a century of controversy resolved by in vivo measurements. *Proc Natl Acad Sci U S A* 111:E5214–E5223. [CrossRef Medline](#)
- Yin C, Shimojo S, Moore C, Engel SA (2002) Dynamic shape integration in extrastriate cortex. *Curr Biol* 12:1379–1385. [CrossRef Medline](#)
- Zimmermann E, Morrone MC, Burr D (2012) Visual motion distorts visual and motor space. *J Vis* 12:10. [CrossRef Medline](#)