

Moving a Shape behind a Slit: Partial Shape Representations in Inferior Temporal Cortex

Anna Bognár^{1,2} and Rufin Vogels^{1,2}

¹Laboratorium voor Neuro- en Psychofysiologie, Department of Neurosciences, KU Leuven, Leuven, 3000, Belgium, and ²Leuven Brain Institute, KU Leuven, Leuven, 3000, Belgium

Current models of object recognition are based on spatial representations build from object features that are simultaneously present in the retinal image. However, one can recognize an object when it moves behind a static occlude, and only a small fragment of its shape is visible through a slit at a given moment in time. Such anorthoscopic perception requires spatiotemporal integration of the successively presented shape parts during slit-viewing. Human fMRI studies suggested that ventral visual stream areas represent whole shapes formed through temporal integration during anorthoscopic perception. To examine the time course of shape-selective responses during slit-viewing, we recorded the responses of single inferior temporal (IT) neurons of rhesus monkeys to moving shapes that were only partially visible through a static narrow slit. The IT neurons signaled shape identity by their response when that was cumulated across the duration of the shape presentation. Their shape preference during slit-viewing equaled that for static, whole-shape presentations. However, when analyzing their responses at a finer time scale, we showed that the IT neurons responded to particular shape fragments that were revealed by the slit. We found no evidence for temporal integration of slit-views that result in a whole-shape representation, even when the monkey was matching slit-views of a shape to static whole-shape presentations. These data suggest that, although the temporally integrated response of macaque IT neurons can signal shape identity in slit-viewing conditions, the spatiotemporal integration needed for the formation of a whole-shape percept occurs in other areas, perhaps downstream to IT.

Key words: body patch; inferior temporal; macaque; object recognition; slit-viewing; visual occlusion

Significance Statement

One recognizes an object when it moves behind a static occluder and only a small fragment of its shape is visible through a static slit at a given moment in time. Such anorthoscopic perception requires spatiotemporal integration of the successively presented partial shape parts. Human fMRI studies suggested that ventral visual stream areas represent shapes formed through temporal integration. We recorded the responses of inferior temporal (IT) cortical neurons of macaques during slit-viewing conditions. Although the temporally summated response of macaque IT neurons could signal shape identity under slit-viewing conditions, we found no evidence for a whole-shape representation using analyses at a finer time scale. Thus, the spatiotemporal integration needed for anorthoscopic perception does not occur within IT.

Introduction

Current models of the ventral visual stream (Kriegeskorte, 2015; Yamins and DiCarlo, 2016; Kar et al., 2019) are based on spatial representations of an object's image. For instance, the activation of units of convolutional neural network models depends on the

spatial integration of local stimulus features. However, one can recognize an object when it moves behind a static occlude, and only a small part of its shape is visible at a given moment in time (i.e., seeing a dog walk behind a slightly open door) (Parks, 1965; Rock and Sigman, 1973). Psychophysical studies showed that the perception of a complete moving figure when it only is revealed, one tiny part at a time, through a static slit (anorthoscopic perception [AP]) is not because of retinal smearing because of pursuit eye movements but reflects temporal integration of spatially fragmented shape information (McCloskey and Watkins, 1978; Morgan et al., 1982). AP provides a challenge to current spatial-based models of object recognition since successive shape elements stimulating the same retinal strip must be integrated over time to obtain a representation of the object.

Human fMRI studies reported activations in dorsal and ventral visual stream areas during AP (Yin et al., 2002; Reichert et

Received Feb. 15, 2021; revised June 4, 2021; accepted June 8, 2021.

Author contributions: R.V. and A.B. designed research; R.V. and A.B. edited the paper; R.V. and A.B. wrote the paper; A.B. performed research; A.B. contributed unpublished reagents/analytic tools; A.B. analyzed data.

This work was supported by Fonds voor Wetenschappelijk Onderzoek Vlaanderen (Odysseus G0007.12), KU Leuven Grant C14/17/109, and European Research Council, European Union's Horizon 2020 research and innovation program (Grant Agreement 856495). We thank P. Kayenbergh, I. Puttemans, A. Hermans, G. Meulemans, W. Depuydt, C. Ullens, S. Verstraeten, and M. Depaep for technical and administrative support.

The authors declare no competing financial interests.

Correspondence should be addressed to Rufin Vogels at rufin.vogels@kuleuven.be.

<https://doi.org/10.1523/JNEUROSCI.0348-21.2021>

Copyright © 2021 the authors

al., 2014; Orlov and Zohary, 2018). In particular, these fMRI studies suggested that, among other areas (e.g., hMT/V5), the lateral occipital complex (LOC) area, a key area of the human ventral visual stream, is more active when through a narrow slit a whole object is perceived than when it is seen as isolated shape fragments. One recent fMRI study (Orlov and Zohary, 2018) presented nonfamiliar shapes that moved behind either a vertically or horizontally oriented narrow slit. Using multivoxel pattern analysis, they found that the pattern of activation in LOC encoded the shape of the object when it was moving behind the narrow slit. This suggested to the authors that LOC represents a whole-shape percept based on the temporal integration of the slit-views.

These intriguing fMRI data cannot answer the question of how LOC neurons respond in the slit-viewing condition because of the limited temporal resolution of fMRI. If the neurons perform a temporal integration of the partially occluded moving shapes, the response selectivity should increase over successive views. On the other hand, the hemodynamic response in the slit-viewing condition might result from different neurons, each responding to different shape features that are revealed during a few consecutive slit-views. If fragments of different shapes excite different neurons, then the pattern of activity across neurons, temporally integrated by the slow hemodynamic response function (HRF), will differ among shapes, yielding decoding of shape identity from the slit-viewing condition. Orlov and Zohary (2018) also used a condition in which the slit-views were presented in random order; and for this condition, no significant shape decoding was possible in LOC. However, the 60 Hz frame rate in that random condition will have produced flicker and strong forward-backward masking of individual slit-views. Hence, dissociation of slit-viewing representations of shape fragments from temporally integrated whole-shape representations is difficult with that control.

To examine the time course of the responses during AP, we recorded single neurons of the macaque inferior temporal (IT) cortex during slit-viewing of shapes. Macaque IT is assumed to be the homolog of human LOC (Denys et al., 2004). To increase the effectiveness of the shapes in driving selective activity, we performed recordings in and close to an fMRI-defined body patch of the anterior superior temporal sulcus (ASB) (Kumar et al., 2019), using silhouettes of animals, known to produce strong selective responses in that patch (Bao et al., 2020). The shapes were presented statically or when moving behind a static narrow vertical or horizontal slit. Analysis of single-unit and population responses showed that the IT neurons responded selectively to shape fragments during slit-viewing but did not temporally integrate the shape, even when the monkey was matching static whole shapes to shapes presented during slit-viewing.

Materials and Methods

Subjects

Three male rhesus monkeys (*Macaca mulatta*; MG, MB, and MT) were implanted with an MR-compatible headpost and a recording chamber targeting ASB, using surgical procedures under full anesthesia as described previously (Popivanov et al., 2014). Animal care and experimental procedures complied with the National, European, and National Institute of Health guidelines and were approved by the Ethical Committee of the KU Leuven Medical School.

fMRI body patch localizer

The monkeys were scanned on a 3T Siemens Trio scanner following published standard procedures (Vanduffel et al., 2001). Functional MR images were acquired using a custom-made 8-channel monkey coil

(Ekstrom et al., 2008) and a gradient-echo single-shot EPI sequence (for more details, see Popivanov et al., 2012). In Monkeys MB and MT, we used the block design procedure of Popivanov et al. (2012), showing 20 images of monkey bodies, monkey faces, objects, mammals, birds, and fruits/vegetables while the subjects were performing a passive fixation (PF) task for a juice reward. ASB is the most anterior body patch in the superior temporal sulcus, defined by the contrast monkey bodies minus objects (for further details, see Kumar et al., 2019). The fMRI data of Monkey MG were obtained in the context of previous studies (Taubert et al., 2015; Vinken et al., 2018). This monkey was scanned during PF with stimuli of different classes that were identical to those used by Tsao et al. (2003). The contrast to define ASB was bodies (without heads) minus faces, fruits, tools, and hands. The fMRI maps were coregistered with an anatomic MRI of each monkey, and these images were used to position recording chambers and guide tube locations. Further details about the procedure used to target fMRI-defined body patches can be found in our previous publications (Popivanov et al., 2014; Kumar et al., 2019).

Electrophysiological recordings

Standard single-unit recordings were performed with epoxylite-insulated tungsten microelectrodes (FHC; *in situ* measured impedance of $\sim 1\text{ M}\Omega$) using techniques as described previously (Sawamura et al., 2006). Briefly, the electrode was lowered with a Narishige microdrive into the brain using a stainless-steel guide tube that was fixed in a standard Crist grid positioned within the recording chamber. After amplification and filtering between 540 Hz and 6 kHz, spikes of a single unit were isolated online using a custom amplitude- and time-based discriminator. The recording grid locations were defined so that the electrode targeted ASB or neighboring sites in the left hemisphere. We used the body patch localizer to increase the frequency of finding neurons that responded strongly and selectively to the used shapes. Previous studies showed that ASB neurons respond well to images of four-legged animals (Kumar et al., 2019; Bao et al., 2020) and ASB is strongly activated by silhouettes of animals (Bao et al., 2020). Thus, by using silhouettes of animals as shapes and by targeting ASB and surrounding sites as recording locations, we aimed to increase the efficiency of the recordings.

The position of one eye was continuously tracked using an infrared video-based tracking system (SR Research EyeLink; sampling rate 1 kHz). Stimuli were displayed on a CRT display (Philips Brilliance 202 P4; 1024 \times 768 screen resolution; 75 Hz vertical refresh rate) at a distance of 57 cm from the monkey's eyes. The onset and offset of the stimuli were signaled using a photodiode, detecting luminance changes of a small square in the corner of the display (but invisible to the monkey), placed in the same frame as the stimulus events. A digital signal processing-based computer system developed in-house controlled stimulus presentation, event timing, and juice delivery while sampling the photodiode signal, vertical and horizontal eye positions, spikes, and behavioral events. Time stamps of the recorded spikes, eye positions, bandpass-filtered electrode signal (sampling rate 40 kHz), stimulus, and behavioral events were stored for offline analyses. Isolation of the single units was checked offline using the spike-sorting software of the Spike2 analysis package.

Stimuli and tasks

Silhouettes. For the single-unit recordings in the PF task (Monkeys MG, MT, and MB), the stimulus set consisted of 70 black silhouettes of animals (for examples, see Fig. 1A). The shapes were presented on a gray background. Both the vertical and horizontal extent of the shapes was fixed to 4.8° (i.e., their bounding box was a square). The equal shape size ensured that, for moving shapes behind a slit, the duration during which the shape fragments were visible was constant among all the shapes and motion directions. For the behavioral training and subsequent testing (Monkey MG), we introduced a new stimulus set of 90 silhouettes of animals, all having equal size.

Search test. The trial started with the onset of a small red square (size 0.2°) on top of a 15° sized square that consisted of visual noise, presented on the gray background of the display. The static noise was created by randomly positioning white and black pixels ("salt and pepper" noise;

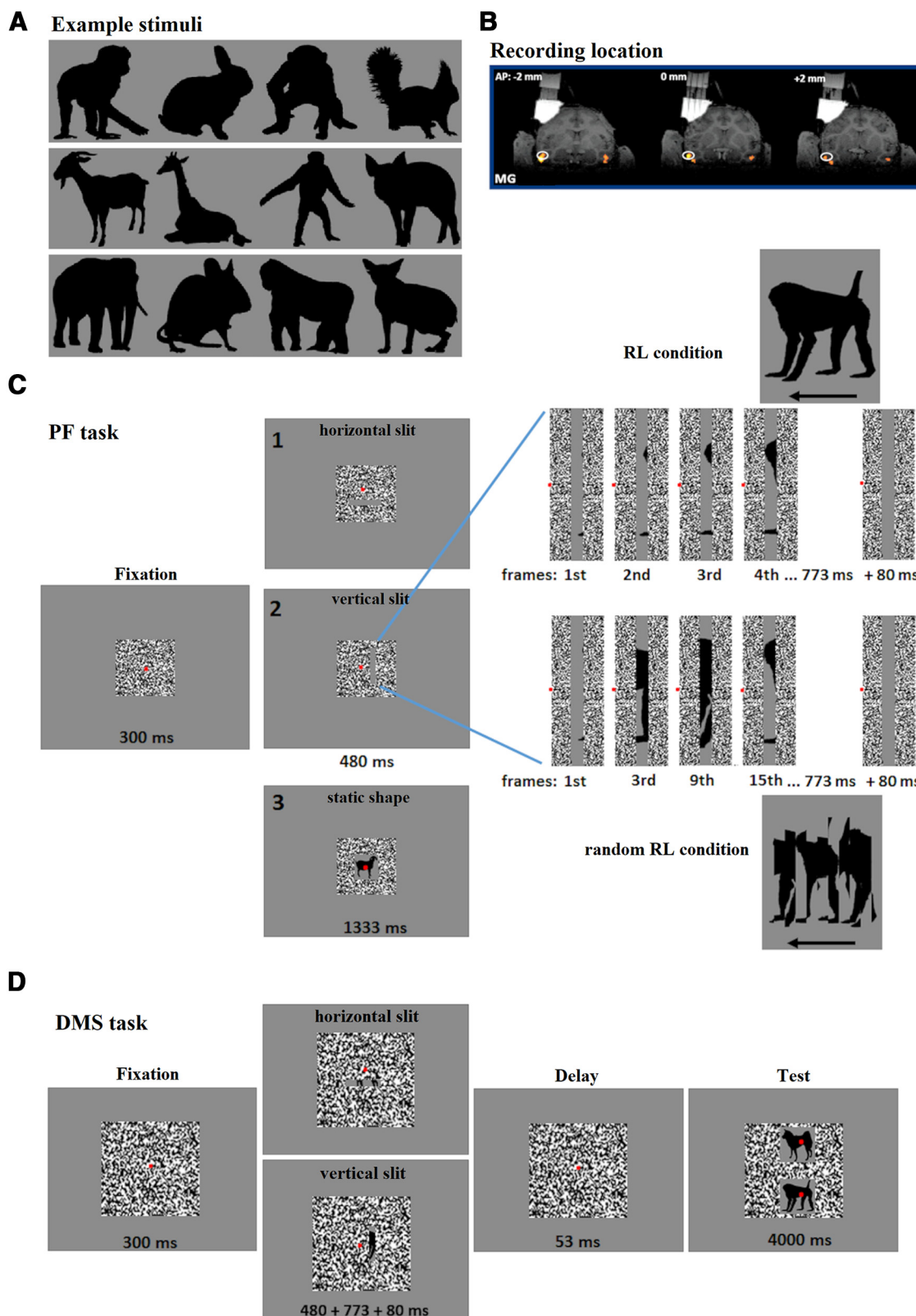


Figure 1. Stimuli and tasks. **A**, Examples of shapes used in the electrophysiological experiments. **B**, Coronal MRI sections showing the location of the body patch ASB (indicated by the oval outline) in Monkey MG. The fMRI activations of the body patch localizer are shown in hot colors (threshold $t > 5$). **C**, Slit-viewing test during PF. Following a 300 ms fixation period, an empty horizontal gray slit (**C1**), a vertical gray slit (**C2**), or a gray square aperture that included a static shape (**C3**) was presented in the noise background. The static shape configuration was presented for 1333 ms, equal to the duration of the slit. After 480 ms of empty slit presentation, shape passed behind the slit. The movie displaying the shape fragments lasted 773 ms. After the presentation of the shape, the empty slit remained present for another 80 ms. A red fixation target was present during the entire trial on top of the stimuli. Its continuous fixation resulted in a fluid reward at the end of the trial. The shape fragments, presented in the first 4 frames (75 Hz frame rate) of a vertical slit condition in which the shape moved leftwards (RL condition) are shown to the right. Below it, we show shape fragments of the corresponding “random” condition. In that condition, each of 9 shape fragments was presented for 6 frames (80 ms), except the first and the last one, which were presented for 2 frames. Below the slit-view examples of the random condition, the spatially concatenated shape parts that were presented during this

pixel size = 0.03°). The monkeys had to fixate the red target for 250 ms. Then, the shape was presented for 350 ms on the noise background, centered behind the fixation target. The monkey had to continue fixation during the stimulus presentation and for a period of 108 ms after stimulus offset. Continuous fixation within a fixation window of $2^\circ \times 2^\circ$ was rewarded with a drop of juice. The 70 shapes were shown in a pseudo-random order for at least five unexpired trials each. All neurons were tested using this procedure, and the neuron was examined in further tests when a response was notable in the online peristimulus time histograms (PSTHs) for at least one of the shapes. Based on the responses to the individual shapes, we selected a shape that produced the highest response (“best”), a shape for which there was no or a weak response (“worst” shape), and, for most recordings, also a shape that produced a response intermediate (“medium”) between the best and worst shape.

Slit-viewing test during PF. The shapes that were presented in this test were those selected during the preceding search test for the neuron under investigation. These shapes were presented under three conditions (see Fig. 1C): (1) static shape, (2) slit-viewing with the original shape, and (3) slit-viewing with randomly ordered shape parts. The trial started with the onset of the fixation target on the top of the previously described static noise background. Following a 300 ms fixation period, either an empty gray slit (width = 0.48°; length = 7.2°) or a gray square aperture (size = 5.3°) that included a static shape was presented in the noise background (see Fig. 1C). The static shape configuration was presented for 1333 ms, equal to the duration of the slit. In the slit-viewing conditions, the shape became visible in the slit aperture after 480 ms of empty slit presentation. The movie displaying the shape fragments lasted 773 ms (see Fig. 1C). After the presentation of the successive shape parts, the empty slit remained present for another 80 ms. The fixation target was present during the entire trial on top of the stimuli, and its continuous fixation (fixation window size = $2^\circ \times 2^\circ$) resulted in a fluid reward at the end of the trial.

The shape fragments shown in the slit depended on the condition. In the slit-viewing with original shape conditions (“original” condition), one of the selected shapes was moving smoothly behind the slit with a speed of 6.2°/s. The slit was oriented either vertical or horizontal and presented at an eccentricity of 2° in the contralateral visual field (centered on the horizontal meridian) or below the fixation target (centered on the vertical meridian), respectively. In the case of the vertically oriented slit, the shape was either moving leftward or rightward. In the case of the horizontally oriented slit, the shape was moving upward or downward.

The speed and duration of the motion were highly similar to that used in a previous human fMRI study (Orlov and Zohary, 2018). These authors did not use a noise background, but we included it to have additional occlusion cues. In our conditions, the shape is perceived as moving behind the noise background, becoming partially visible through the static slit aperture. Another difference between our display and that of the human fMRI study (Orlov and Zohary, 2018) is that we had separate trials for the two motion directions, instead of a presentation of the two motion directions immediately after each other in a single trial. Unlike in the human fMRI study (Orlov and Zohary, 2018), in which the whole shapes were presented only after the slit-viewing conditions, we

presented the whole shapes in the search task before the slit-viewing tasks and interleaved them with the slit-viewing conditions. This could only have increased the percept of the shape during slit-viewing. The presentation of the empty slit, well before the shape became visible, aimed to reduce potential neural responses to the onset of the slit itself by the time of shape onset. The Orlov and Zohary (2018) study used novel, unfamiliar shapes, while our shapes were familiar to the monkey since these were presented repeatedly while searching for neurons. However, which shape was presented during slit-viewing on a particular trial was unpredictable.

In the slit-viewing with randomly ordered shape fragments conditions (“random” condition), we presented frames from the slit-viewing with original shape conditions in a pseudo-random order (see Fig. 1C). The order of the shape parts was random, except that fragments that followed each other in the original slit-viewing condition were not allowed to be sequential in the random condition. However, we made two changes to the random-order conditions used in a previous human fMRI study (Orlov and Zohary, 2018). First, we presented only frames with nonoverlapping shape segments, and these were presented for 6 successive frames (80 ms) each. The latter reduced the contribution of forward and backward masking (Kovacs et al., 1995a) to the responses to the partial shape views in these random conditions, which are expected to have a strong impact when using a random ordering of the frames at the original frame rate, as in the human fMRI study (Orlov and Zohary, 2018). Thus, we presented 9 fragments in random order for 6 frames each, and these were preceded and followed by two frames of the start and end fragments, respectively, of the corresponding slit-viewing with original shape condition (see Fig. 1C). The total duration equaled that of the original slit-viewing condition. Second, the order of the frames was fixed across trials, allowing us to average across trials the responses using short time bins. The shape fragments of the random condition are a subset of those shown in the single frames of the original slit-viewing condition since the partial shape sections of successive frames in the latter condition partially overlapped because of the smooth motion of the shape. To control for the partial overlap of the shape fragments in the original slit-viewing and the random condition, we also made recordings in a subset of neurons with a third kind of display in which the same fragments that were presented in the random condition were shown in their correct order. This yielded a similar percept as in the corresponding original slit-viewing conditions, except for some minor jumps across the 6 frames long presentations. We denote these displays as “jumping” displays. Thus, in the random and jumping conditions, the same shape fragments were presented through the aperture but in a different order.

The different conditions were presented interleaved in random order for at least 10 trials each.

Snapshot test. In this test, we measured the responses to static presentations of individual shape fragments. Thus, we presented the 11 slit-view displays (shape part together with background noise pattern) of the random (and jumping) conditions separately in different trials with an intertrial interval of at least 133 ms. During a trial, the monkey was required to fixate for 300 ms, followed by a presentation of the empty slit for 480 ms, after which the shape fragment was displayed for 80 ms. Then, the monkey needed to continue fixation for another 300 ms to obtain the juice reward. The individual snapshots were presented in a randomly interleaved fashion for at least 5 trials each. The snapshot test was preceded by the search task to select two shapes with high responses.

Delayed matching to sample (DMS) test. This test was used in Monkey MG for both behavior and single-unit recordings (see Fig. 1D). A trial started with the onset of the fixation target on top of a square noise background (size = 20°). After 300 ms of fixation, the sample stimulus was presented. This could be a slit-viewing movie with identical parameters as in the PF task. Following the last shape fragment, the empty slit remained present for 80 ms. The sample stimulus could also be a presentation of the whole shape (duration 1333 ms) as in the PF test. Then the background noise pattern without slit was presented for 53 ms after which two shapes were presented. The shapes were shown above and below the fixation target at an eccentricity of 4.6° . One of the two shapes corresponded to the one shown as sample stimulus, and the monkey was

←

random condition are shown with the first presented fragment to the left. **D**, DMS test. After 300 ms of fixation of the red fixation target, on top of a square noise background, the sample stimulus was presented. This could be a slit-viewing movie with identical parameters as in the PF task: 480 ms of empty slit presentation, followed by a display of the shape fragments for 773 ms and the empty slit for 80 ms. An actual shape fragment of a single frame is shown for the horizontal and vertical slit presentation. The sample stimulus could also be a presentation of the whole shape (not illustrated). Then, the background noise pattern without slit was presented for 53 ms after which two shapes were presented simultaneously. The shapes were presented above and below the fixation target. One of the two shapes corresponded to the one shown as sample stimulus, and the monkey was required to make a saccade to the stimulus that matched the sample stimulus. The match and nonmatch stimuli stayed on the screen until the monkey made a saccade to one of the stimuli or for maximally 4 s. Only correct responses were rewarded with a fluid reward.

required to make a saccadic eye movement to the shape that matched the sample stimulus. The match and nonmatch stimuli stayed on the screen for 4 s or until the monkey made a saccade to one of the stimuli. Correct responses were rewarded with a fluid reward.

The monkey was trained extensively in this task with a large variety of 60 shapes. After this training, we presented old and novel shapes (30) as sample stimuli. The match and nonmatch stimuli could be novel, old and novel, or both old. Using a subset of stimuli (10 old, 10 novel shapes), sample stimuli could be slit-view presentations of either moving shapes or randomly ordered snapshots of the slit-views as in the random PF condition. The different tested combinations of sample, match, and nonmatch stimuli will be described in Results.

After the behavioral testing, we recorded single neurons when the monkey was performing the DMS task. For each neuron, we selected three shapes using the search test. The three shapes were presented either as static whole shapes or during slit-viewing. The nonmatch stimuli were other shapes from the search test. To assess whether the execution of the task or the extensive training influenced the responses of the neuron, we also recorded neurons using the PF test. The latter recording phase (post-DMS) was performed after finishing the recording period during which the animal performed the DMS task.

Data analysis

Responsiveness and selectivity. Because we searched for neurons using static, whole shapes with the search test, all neurons responded well to at least the best static shape condition of the subsequent slit-viewing test. We assessed significant responses of each neuron to the original slit-viewing conditions of the slit-viewing test using a three-way Split-Plot ANOVA with the repeated-measures factor “epoch” (9 levels corresponding to 9 windows of 100 ms each, starting 100 ms before motion onset) and between-trial factors “shape” (three levels: best, medium, and worst) and “motion direction” (four levels: leftward [RL], rightward [LR], upward [DU], and downward [UD]). We used windows of 100 ms for the factor “epoch” since we noted during the recordings that the neurons responded during a limited period of the movie, and we wished to capture such modulation of the response during the slit-viewing movie. Neurons that showed a significant effect ($p < 0.05$) of the factor “epoch” or an interaction of the factors “shape” and “epoch” were considered to show a significant response to the slit-viewing stimuli.

For each neuron that showed a significant excitatory response ($<5\%$ of the significant neurons showed inhibition during the slit-viewing period), we computed the mean response, across trials, during the slit-viewing period, using a window of 800 ms that started 50 ms after motion onset. For each monkey and the data pooled across monkeys, we then performed a three-way repeated-measures ANOVA using the responses of each neuron with as repeated factors “shape” (best, medium, worst), “direction” (LR, RL, UD, DU), and “slit-viewing condition” (slit-viewing of the original shape [“original”], slit-viewing with randomly ordered views [“random”]). To assess the significance of the repeated factors and their interaction, we applied sphericity correction using the Greenhouse-Geisser method. A similar analysis was performed for the neurons that were tested with the jumping displays. In that analysis, the factor “slit-viewing condition” had three levels, being “original,” “random,” and “jumping.”

To quantify the extent to which shape and motion direction were encoded in a separable manner, we computed for each responsive neuron a “Separability Index.” This index compares the responses to the 3 shape \times 4 motion direction combinations of the “original” condition to responses predicted under the assumption that the response to each combination results from independent tuning along the shape and motion direction dimensions. We followed a previously published procedure (Mysore et al., 2010) to compute this index. The mean firing rates, computed in the 800-ms-long window and averaged over trials, for the 12 stimuli were tabulated in a 3×4 response matrix (M) with m and n corresponding to the 3 shapes and the 4 motion directions, respectively. We then computed the singular value decomposition ($M = USV'$) of the response matrix. The predicted response was the product of the first columns of U and V of the singular value decomposition. The

Separability Index equals the squared Pearson correlation (r^2) between the actual and predicted responses.

We used two motion axes, horizontal and vertical, and two motion directions for each axis. To quantify the effect of motion direction, within and across axes, we computed “Direction Indices.” We took the responses (800 ms analysis window) to the best shape in the “original” slit-viewing condition and determined the best motion direction using the mean responses computed over half of the trials. Then, for the responses in the remaining half of the trials, we computed the Direction index as follows:

$$\text{Direction Index} = (\text{Response best direction} - \text{Response direction } i) / (\text{Response best direction} + \text{Response direction } i),$$

with motion direction i being either the direction opposite to the best direction along the same axis or the two directions of the orthogonal axis. Thus, we computed three Direction Indices for each responsive neuron: one for opposite directions along the same axis and two for directions along the axis orthogonal to the “best” axis. The index can range in principle from -1 to 1 , but negative values indicate conceptually an absence of direction selectivity (as a value of zero), since responses higher than those to the best direction, measured in other trials, will reflect noise (trial-to-trial variability).

Time course and responses to shape-fragments. The neurons responded only during a limited phase of the slit-viewing movie. To quantify the breadth of the response phase of a single neuron, we estimated the duration of the response at half-height. This estimation was performed for the “original” slit-viewing condition that produced the maximum response. First, to reduce noise, we smoothed the mean response (bin of 1 ms), averaged across trials, using a Gaussian kernel with an SD of 10 ms. Then, we defined the “peak duration” as the period during which the smoothed response was at least half the smoothed peak response. We used the peak duration metric to compare the duration of the response phase among monkeys and between the PF and DMS tasks in Monkey MG.

To uncover the slit-views to which the neuron responded in the “original” slit-viewing condition, we applied for each motion direction the following procedure, akin to reverse correlation. We binned the mean responses in bins of 75 ms and then assigned each binned response to the shape part that was presented 70 ms before the start of that bin. Doing so, we obtained a vector of the responses to the shape fragments during slit-viewing and that for each of the four directions. These vectors were then visualized on an image depicting the spatially concatenated slit-views. Further quantification was accomplished by binning the elements of the vector in 11 bins, corresponding to the shape fragments that were presented in the snapshot test.

The responses obtained in the snapshot test were computed using an analysis window of 350 ms that started 50 ms after stimulus onset. The responses to the 11 shape fragments obtained from the snapshot test were then correlated with the responses to the same fragments as estimated from the slit-viewing presentation (see reverse correlation procedure above).

Decoding of shape identity. We decoded the shape identity from the responses of the neurons in the “original” slit-viewing and static presentation conditions. We performed the decoding on the data pooled across monkeys. For decoding, we used the Neural Decoding Toolbox (Meyers, 2013) and linear Support Vector Machines as classifiers. For each neuron, 10 trials of the 15 conditions (original slit-viewing conditions [12] and the responses for static stimuli [best, medium, and worst]) were used in the analysis. We made pseudo-population responses by concatenating single-trial responses of the successively recorded neurons in a vector. Thus, each vector represented the response of the population of neurons on a trial. The responses of each neuron were z -normalized across stimulus conditions so that each neuron contributed equally. The classifier was trained using fivefold cross-validation to control for overfitting. The reported classification accuracies are all based on zero-one loss function results. SDs of classification scores were calculated across 50 cross-validated resamplings of the pseudo-population vectors. To test

whether the decoding results were above chance, a permutation test (1000 permutations) with shuffled condition labels was used.

We performed two types of decoding analyses: one using the response averaged in the 800-ms-long window and a second one using shorter 100 ms bins. In both analyses, we trained the classifier for one stimulus condition (e.g., slit-viewing LR) and then tested the classification accuracy of that classifier for the independent test trials of that condition and the other conditions (e.g., static whole shape, RL, UD, DU). The latter tested whether shape classification tolerated a change of the viewing condition. In the case of the short bin decodings, we trained and tested the classifier for all possible combinations of training and testing bins, ranging from –200 until 1000 ms relative to motion, or static stimulus onset. This analysis allowed an assessment of the temporal specificity of shape encoding during slit-viewing.

Behavioral performance in DMS task. We included only unaborted trials in which the monkey made a saccade to one of the two test stimuli. Percent correct responses were computed for all trials of a test condition, as will be specified in Results. CIs (95%) of percent correct were computed using the Binomial distribution: <https://www.graphpad.com/quickcalcs/confinterval1/>.

Eye movements. Eye positions along the horizontal and vertical dimensions were analyzed separately for each of the motion directions during slit-viewing. Before averaging, we subtracted for each trial the mean eye position in a 200 ms period before motion onset from the eye positions measured after motion onset. For each monkey, we averaged the baseline-subtracted eye positions per shape and motion direction for the original slit-viewing conditions. CIs (95%) were computed for each time point with bootstrap resampling.

Experimental design and statistical analysis. We used both parametric (ANOVA) and nonparametric tests. The factors and design of the ANOVA are described above and in the corresponding Results sections. Parametric tests were used only when no nonparametric, distribution-free tests were available. ANOVAs were performed using the R statistical software package and nonparametric tests, except noted otherwise, were performed using MATLAB functions.

Results

We examined the responses of IT neurons to silhouettes of animals that were moving behind a static narrow slit in an opaque occluder (Fig. 1). The 0.48° wide slit was presented at 2° eccentricity, to avoid smooth pursuit of the moving shape fragments. It was oriented either horizontal or vertical, and only 10% of the shape was visible during a single frame of the movie. Initially, we recorded the responses of well-isolated single units during slit-viewing when 3 monkeys were performing a PF task. After this series of recordings, we trained 1 monkey in a DMS task and assessed whether he was able to match the partial views of a moving shape, passed behind the slit, with the static unoccluded presentation of the same shape. We also recorded responses of single units of the same patch when the monkey was performing the DMS task using slit views as sample stimuli. After these recordings, we again measured the responses of single neurons during slit-viewing in the PF task in the same monkey that was tested in the DMS task.

Responses and selectivity in PF task

We examined the responses of single body patch neurons (ASB; Fig. 1B) to slit-viewing using three equally sized silhouette shapes of animals. For every single neuron, we selected the three shapes using the responses to 70 shapes that were presented in a search test (see Materials and Methods). One of them, labeled “best,” produced the largest response of the 70 shapes, a second one, the “worst” shape, no or the weakest response, and the third shape, the “medium” one, a response in-between the best and worst shape. These three shapes were presented when moving behind

the narrow slit in either one of two directions for each of 2 slit orientations (Fig. 1) during PF. In the same test, we also presented the same three shapes without motion in a large aperture of the occluding surface on a gray background.

We recorded the responses of 196 IT neurons, responsive to static whole shapes, in the slit-viewing test. We assessed for each neuron whether it responded significantly in at least one of the slit-viewing conditions with a three-way Split-Plot ANOVA with a repeated-measures factor “epoch” (9 levels corresponding to 9 windows of 100 ms each, starting 100 ms before motion onset; see Materials and Methods) and between-trial factors “shape” (best, medium, and worst) and “motion direction.” The very large majority of the neurons showed a significant effect of the factor “epoch” in each of the 3 monkeys with an excitatory response during slit-viewing [MG: 87% ($N=108$); MT: 100% ($N=63$); MB: 100% ($N=25$)]. Also, the responses of most of these neurons were modulated by shape [MG: 63% ($N=108$); MT: 89% ($N=63$); MB: 72% ($N=25$)] or motion direction [(MG: 69% ($N=98$); MT: 89% ($N=63$); MB: 80% ($N=25$))].

Figure 2 shows the responses of a responsive single neuron to the slit-viewing conditions and the static shape presentations. As expected from the preceding Search test, the neuron produced the largest response to the selected best static shape, no response to the selected worst static shape, and an intermediate response to the medium static shape. In the slit-viewing conditions, this neuron did not respond to the slit-onset itself, which occurred 480 ms before the shape started to move. The neuron responded when the best shape was moving along the horizontal axis behind the slit, while it showed less response when the same shape was moving along the vertical axis. The neuron showed less, if any, response to the medium and worst shapes when these were presented during slit-viewing. Thus, for the horizontal axis slit-viewing conditions, the shape preference fitted that of the whole-shape presentation. The neuron responded only during a brief period of slit-viewing for a particular direction, which was a common finding in our sample of neurons. The timing of this responsive period differed between the two horizontal directions. The parts that were presented at the beginning of one motion direction occurred at the end of the other direction for the same slit orientation. There was no evidence of temporal integration of the responses during the slit-viewing, which was typical for our sample of neurons.

We also tested the same neuron in slit-viewing conditions in which fragments of the same shapes were presented successively but in random order (Fig. 1; see Materials and Methods). This impaired both the perception of smooth motion and the shape. The (random) order of the shape segments was fixed across the trials of the same condition, allowing the computation of PSTHs. The responses in the random slit-viewing conditions are shown in the bluish-shaded panels of Figure 2. This neuron showed overall weaker responses in the random control than in the original slit-viewing conditions, but the strongest response was present for the best shape, horizontal LR random condition.

Other neurons responded with similar peak firing rates in original and random slit-viewing conditions. One example (Fig. 3A) of such neurons responded for the vertical motion slit-viewing conditions of the best and medium shape. It also responded for those directions in the random conditions of the medium (but not best) shape. The neuron of Figure 3B responded in the horizontal slit-viewing conditions of the best shape, and this for both original and random slit-viewing conditions. As for the neuron of Figure 2, both neurons responded during a brief period of the slit-viewing movie, and this for both the original and

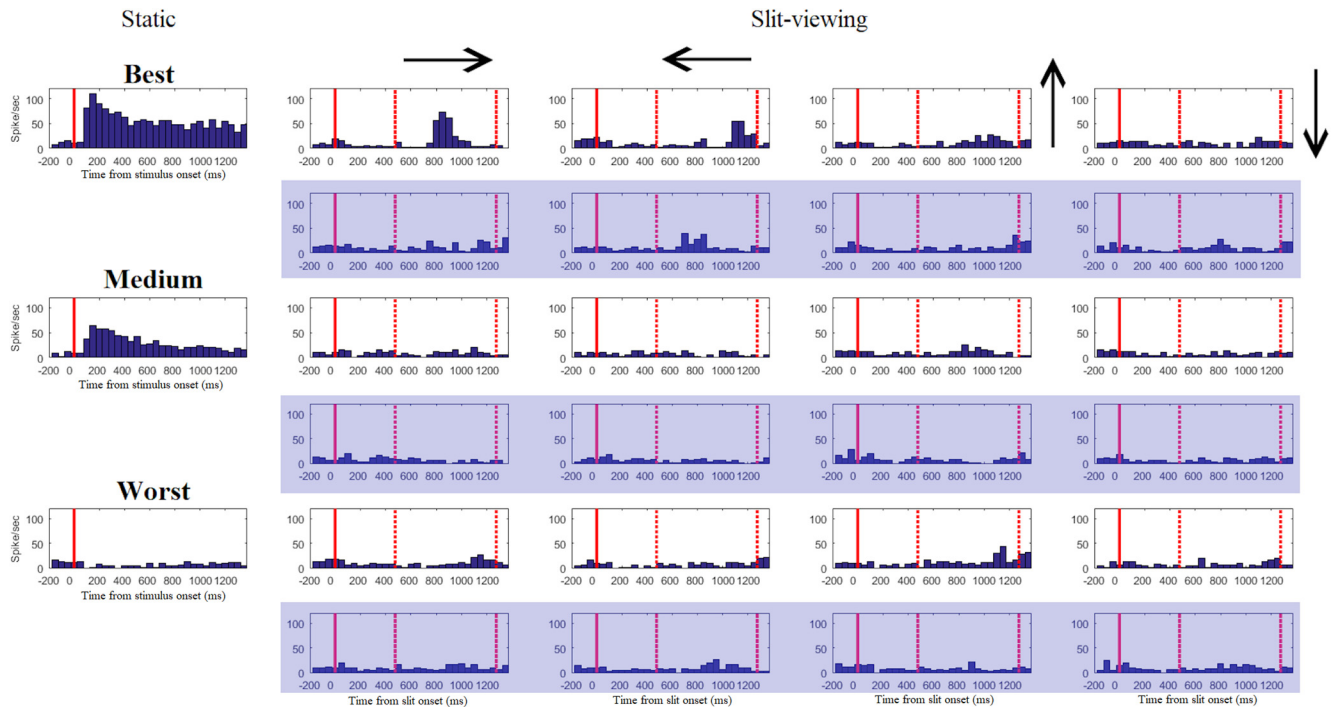


Figure 2. Example neuron recorded in the slit-viewing test during PF. Responses to static whole-shape presentations are shown for the best, medium, and worst shape in the left column PSTHs. Full red line indicates shape onset. The four right columns represent the responses in the slit-viewing conditions. From left to right: rightward motion, leftward motion, upward motion, and downward motion of the shape behind the slit. The first, third, and fifth rows represent the PSTHs for the original slit-viewing conditions for the best, medium, and worst shape, respectively. The second, fourth, and sixth rows (blue shading) correspond to the random slit-viewing conditions. Full red line and 0 indicate the onset of the slit. Stippled red lines indicate the period in which the shape fragments were visible through the slit. Bin width is 50 ms. This neuron was recorded in Monkey MG.

random conditions. Figure 3C shows a neuron that responded somewhat longer during the slit-viewing, but the period during which it responded during slit-viewing depended on motion direction and it showed strong selectivity for the motion axis. None of the neurons in Figure 3 showed evidence of temporal integration of the responses during slit-viewing.

Different neurons responded to different periods of the slit-viewing movies, and those periods also differed between motion directions. Thus, there was substantial heterogeneity among single units of the response profiles for the different slit-viewing conditions. However, when averaging the responses of our sample of responsive neurons ($N = 185$), after normalization of the responses of individual neurons by their maximum firing rate across all conditions (including the static shape and random slit-viewing conditions), we observed a consistent increase in the response shortly after motion onset, which lasted as long as shape fragments were presented in the slit (Fig. 4A). The population responses showed no evidence of temporal integration of the activity during slit-viewing. Indeed, there was no consistent buildup of the response during slit-viewing for the best shape (e.g., LR direction for the best shape). Also, the response dropped to baseline after the last shape part was presented, which conflicts with the hypothetical presence of a whole-shape signal after temporal integration during slit-viewing.

The response during slit-viewing was greater for the best compared with the worst shape, with an intermediate response for the medium shape. The best, medium, and worst shapes were defined based on the response to static presentations of the whole shape. Thus, the shape preference of the population response was invariant to the viewing conditions, although the responses to the static presentations of the whole shapes were markedly greater than the average responses during slit-viewing

of the same shapes (Fig. 4A). The responses for the random conditions tended to be smaller than those for the original slit-viewing conditions; but even for the random conditions, the population responses were larger for the best compared with the worst shape conditions.

To assess the statistical significance of the effect of shape and the difference between random and original slit-viewing conditions, we computed for each neuron the response for each slit-viewing condition (3 shapes \times 4 directions \times random vs original conditions) using an analysis window of 800 ms that started 50 ms after motion onset (the duration of slit-viewing of the shape was 773 ms). We performed a three-way repeated-measures ANOVA of the responses of the 185 neurons with repeated factors shape, motion direction, and original versus random slit-viewing conditions. The factor shape was highly significant ($F_{(1.76,664.5)} = 116.97$; $p = 1.8 \times 10^{-35}$; Greenhouse-Geisser (sphericity)-corrected): the mean response for both the original and random conditions was the largest for the best shape (defined using static whole-shape presentations), intermediate for the medium shape, and the smallest for the worst shape (Fig. 4B). This difference in mean responses across the shape was significant in each monkey. There was a significant effect of motion direction ($F_{(1.5,828)} = 5.13$; $p = 1.2 \times 10^{-2}$; Greenhouse-Geisser-corrected) with on average stronger responses for the horizontal axis (vertical slit) than vertical axis directions (horizontal slit; Fig. 4B). However, this effect was absent in Monkey MT. Furthermore, the factor motion axis is confounded with a difference in visual field location of the slits; hence, this effect is difficult to interpret. Mean responses were significantly greater for the original compared with the random slit-viewing conditions ($F_{(1,184)} = 51.69$; $p = 1.6 \times 10^{-11}$; Fig. 4B), and this effect was significant in each monkey.

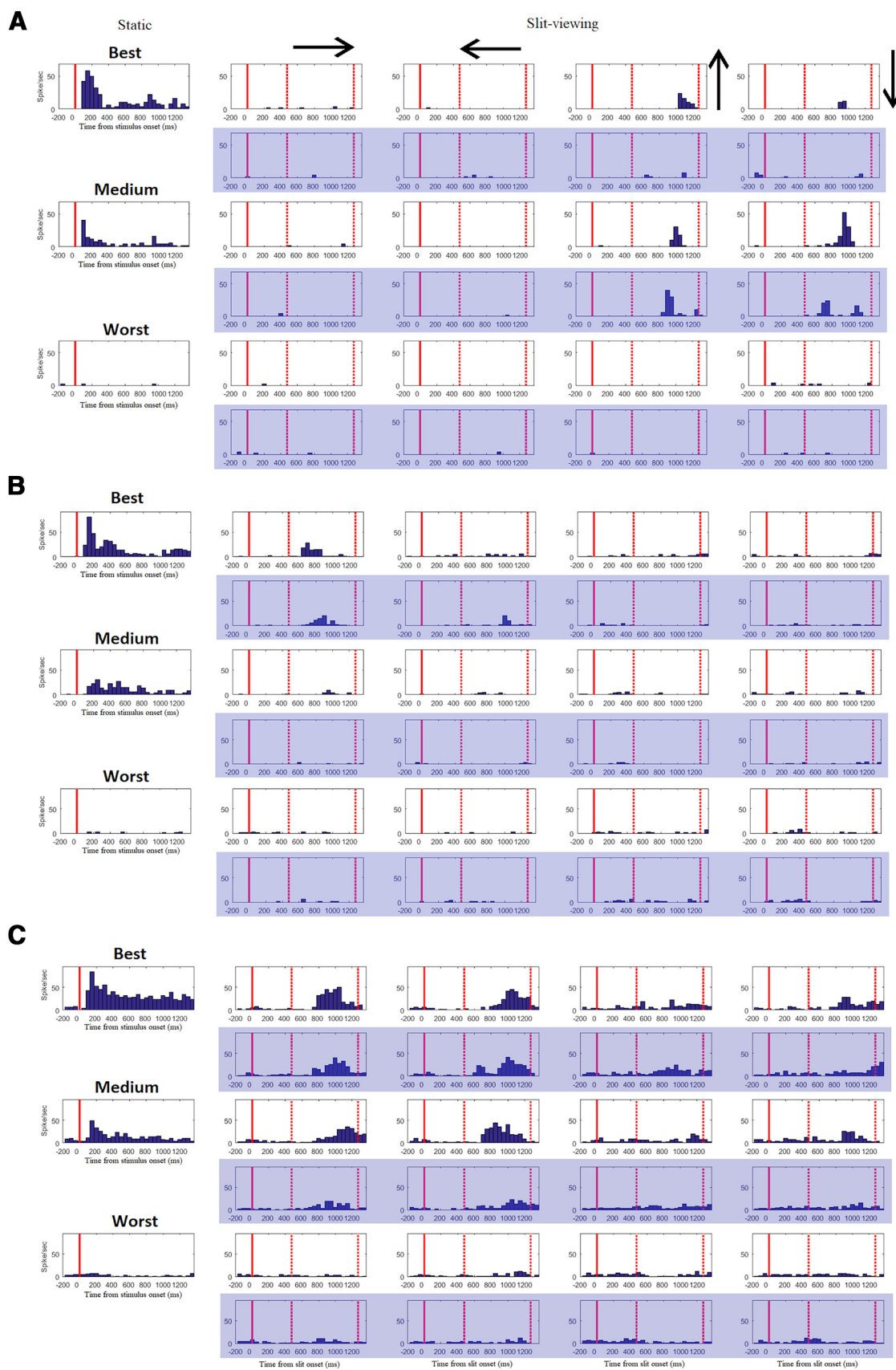


Figure 3. Three example neurons recorded in the slit-viewing test during PF. The neurons were recorded in Monkey MT (**A**), Monkey MB (**B**), and Monkey MG (**C**). Same conventions as in Figure 2.

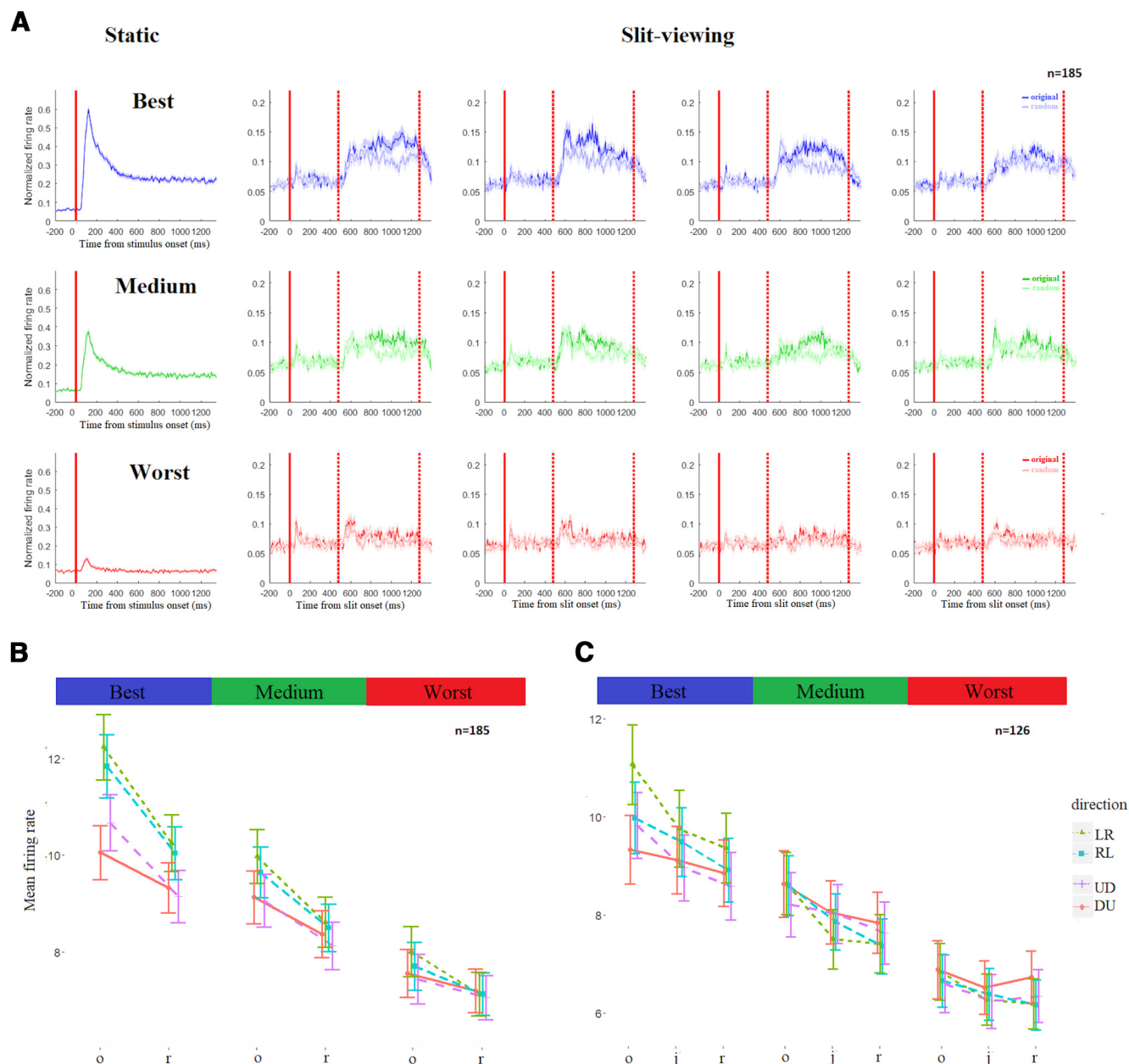


Figure 4. Population responses in the slit-viewing test during PF. **A**, Left column, Population PSTHs of the response to the best (blue), medium (green), and worst (red) static, whole-shape presentations. Other four columns, The responses in the slit-viewing conditions, with motion directions being rightward (LR), leftward (RL), upward (DU), and downward (UD) from left to right (as in Figs. 2 and 3). The responses to the original and random slit-viewing conditions are shown in bold and light blue color, respectively. Bands indicate SEM. We normalized the responses of each neuron by its maximum firing rate (bin width 10 ms) across all conditions. Normalized responses of the 185 neurons of the 3 monkeys were averaged. Note the different y axes for the static and slit-viewing conditions. Other conventions as in Figures 2 and 3. **B**, Mean firing rate of the 185 neurons (3 monkeys) in each of the slit-viewing conditions, for the best, medium, and worst shape. Responses are shown per motion direction and for the original (o) and random (r) slit-viewing condition. Error bars indicate SEM. The mean firing rate (in spikes/s) was computed using an analysis window of 800 ms, starting 50 ms after the shape became visible through the slit. **C**, Mean firing rate of the 126 neurons (3 monkeys) that were tested during PF of the original (o), jumping (j), and random (r) slit-viewing conditions. Responses were sorted per motion direction, shape, and viewing condition. Same conventions as in **B**.

In the original slit-viewing conditions, the visible portions of the shape partially overlapped in successive frames because of the smooth motion of the shape. However, in the random condition, we presented only distinct shape fragments; thus, there was only a partial overlap between the shape parts presented in the original slit-viewing and random conditions. Furthermore, there was no smooth motion in the random condition. To control for these differences, we tested a subset of 126 responsive neurons in the 3 monkeys (MG: $N=38$; MT: $N=63$; MB: $N=25$) with a third set of conditions in which the shape fragments were the same as in the random condition but were shown in the same

order as in the corresponding original condition (“jumping conditions”). As in the random conditions, each shape part was presented for 80 ms in these jumping conditions. The mean responses in the jumping condition were in-between those of the original and random condition (Fig. 4C), and this trend was observed in each monkey. Performing a three-way repeated-measures ANOVA of the responses in the random and jumping conditions of the 126 neurons with repeated factors shape, motion direction, and jumping versus random conditions showed a significant effect of the latter factor ($F_{(1,125)} = 9.08$; $p = 3.12 \times 10^{-3}$). At a superficial level, the greater response to the

jumping compared with the random order condition might be taken as evidence that there is temporal integration of a shape during slit-viewing. However, based on the data we will present below, we prefer an alternative interpretation following the observation that spatially neighboring fragments of a natural shape can contain similar features to which a neuron responds. The latter will result in a longer and stronger response to those successive views, similar to the effect on the response of increasing stimulus duration. When, as in the random condition, these shape features are shown temporally further apart, interleaved with other shape parts, responses are expected to be smaller.

Decoding of shape identity from responses in slit-viewing conditions

A human fMRI study (Orlov and Zohary, 2018) showed that shape identity could be decoded from the multivariate BOLD response in LOC during slit-viewing. Importantly, they reported generalization of classification across slit orientations and for slit-viewing and the static whole-shape presentations. The hemodynamic response is sluggish, integrating neural activity across time. Although single IT neurons showed no evidence of temporal integration of the whole shape during slit-viewing, the shape preference of the average population response during slit-viewing, temporally integrated by averaging in the 800 ms analysis window, was the same as for the static shape presentations (Fig. 4). This suggests that it is possible to decode shape identity from the slit-viewing response and that there is a generalization of shape identity classification across slit-viewing conditions and for the static and slit-view presentations. We examined this by training a linear classifier (Support Vector Machines; see Materials and Methods) to classify shape using as input single-trial pseudo-population response vectors that consisted of the single-trial responses, averaged in the 800 ms window, of the recorded neurons. We trained the classifier using the data of one of the five conditions (static, LR, RL, UD, and DU slit-viewing) and then tested classification for the same, trained, condition (fivefold cross-validation) or the untrained four other conditions (cross-condition classification).

When testing and training were performed using data of the same condition, classification of shape identity was close to or at the ceiling level (Fig. 5A–E, hatched bars) for both static (S) and slit-viewing conditions. This demonstrates that shape identity can be decoded reliably from the temporally integrated responses of the recorded sample of neurons during slit-viewing. For cross-condition classification (Fig. 5A–E, blue bars), the amount of generalization of classification depended on the trained and tested conditions. We observed excellent generalization across orthogonal motion directions of the slit-viewing conditions (e.g., train LR, test RL). The classification scores dropped to ~60% correct but were still significantly above chance (33.3%), when trained and tested motion axes differed (e.g., train LR, test UD), except when training LR and testing DU (Fig. 5B). This demonstrates the generalization of shape classification across slit orientation when integrating the responses across time. Training the classifier with the responses recorded during slit-viewing, we obtained also well above-chance classification of the static shape (Fig. 5B–E). Indeed, the cross-condition test classification scores for the whole shape were similar to those obtained when trained and tested slit orientations were orthogonal. However, training the classifier with the responses to the static whole shape yielded chance classification scores when testing slit-view responses (Fig. 5A). Such asymmetry of cross-condition classification has been observed before when conditions differ markedly in response

strength and signal-to-noise ratio (Van den Hurk and Op de Beeck, 2019), as is also the case here (Fig. 4A). Modeling, using linear Support Vector Machines, has demonstrated that having a low signal-to-noise ratio condition A and a high SNR condition B results in better generalization when testing B after training A than vice versa (Van den Hurk and Op de Beeck, 2019). Also, when only a subset of the informative neurons in condition 2 contains an informative signal for the classifier in condition 1, generalization will be better when training condition 1 and testing 2 than vice versa (Van den Hurk and Op de Beeck, 2019). Also, when only a subset of the informative neurons in Condition 2 contains an informative signal for the classifier in Condition 1, generalization will be better when training Condition 1 and Testing 2 than vice versa (Van den Hurk and Op de Beeck, 2019). Both factors can explain the asymmetry in generalization performance seen in our data since shape selectivity of the single neurons was more robust for the static whole shape presentation compared with the slit-viewing conditions.

In sum, the cross-condition classification data show evidence for generalization across motion direction, slit orientation, and whole-shape versus slit-view presentations when integrating the response throughout the slit-viewing period. These results of monkey IT single-unit data are in line with the generalization of shape identity classification obtained for BOLD activation patterns in human LOC (Orlov and Zohary, 2018). Since the BOLD HRF causes integration of the neural responses across time, it can produce similar results as we obtained here by temporal integration of the spiking activity of single neurons. However, the underlying dynamics of the stimulus representation during slit-viewing are lost when temporally integrating neural responses. To capture the dynamics of the shape representation during slit-viewing, we decoded shape identity with classifiers that were trained and tested with brief 100 ms analysis windows. We performed classification of shape identity for different training and testing periods, and this when trained and tested conditions were identical or differed. Figure 5F shows the classification scores for all possible training-testing time combinations (cross-temporal decoding), starting 200 ms before motion onset and ending 227 ms after motion offset, and this for all the 25 possible combinations of the trained and tested stimulus conditions. When trained and tested conditions were identical (panels along the [top] left to the right diagonal in Fig. 5F), the classification accuracy dropped markedly when training and testing differed by >200 ms for all slit-viewing conditions. This is clearly shown by the reddish left diagonal band in the cross-temporal decoding plots. This temporally specific code contrasted with the more stationary one for the static, whole-shape presentations. Importantly, the cross-temporal decoding plots for opponent trained and tested motion directions showed also a clear diagonal band but from right to left. Thus, the generalization of shape classification across opponent motion directions is highly temporally specific. The mirror symmetry of the cross-temporal decoding plots for identical versus opponent trained and tested motion directions suggests that the neurons responded to shape fragments (“effective shape fragments”) that were visible at a particular moment of the slit-viewing: an effective shape fragment that is visible; for example, at the beginning of the slit-viewing for one motion direction (e.g., LR) will be visible at the end of the slit-viewing for the opponent motion direction (e.g., RL). When trained and tested conditions consisted of different motion axes (horizontal vs vertical), there will not be a correlation of effective shape fragments across time between the

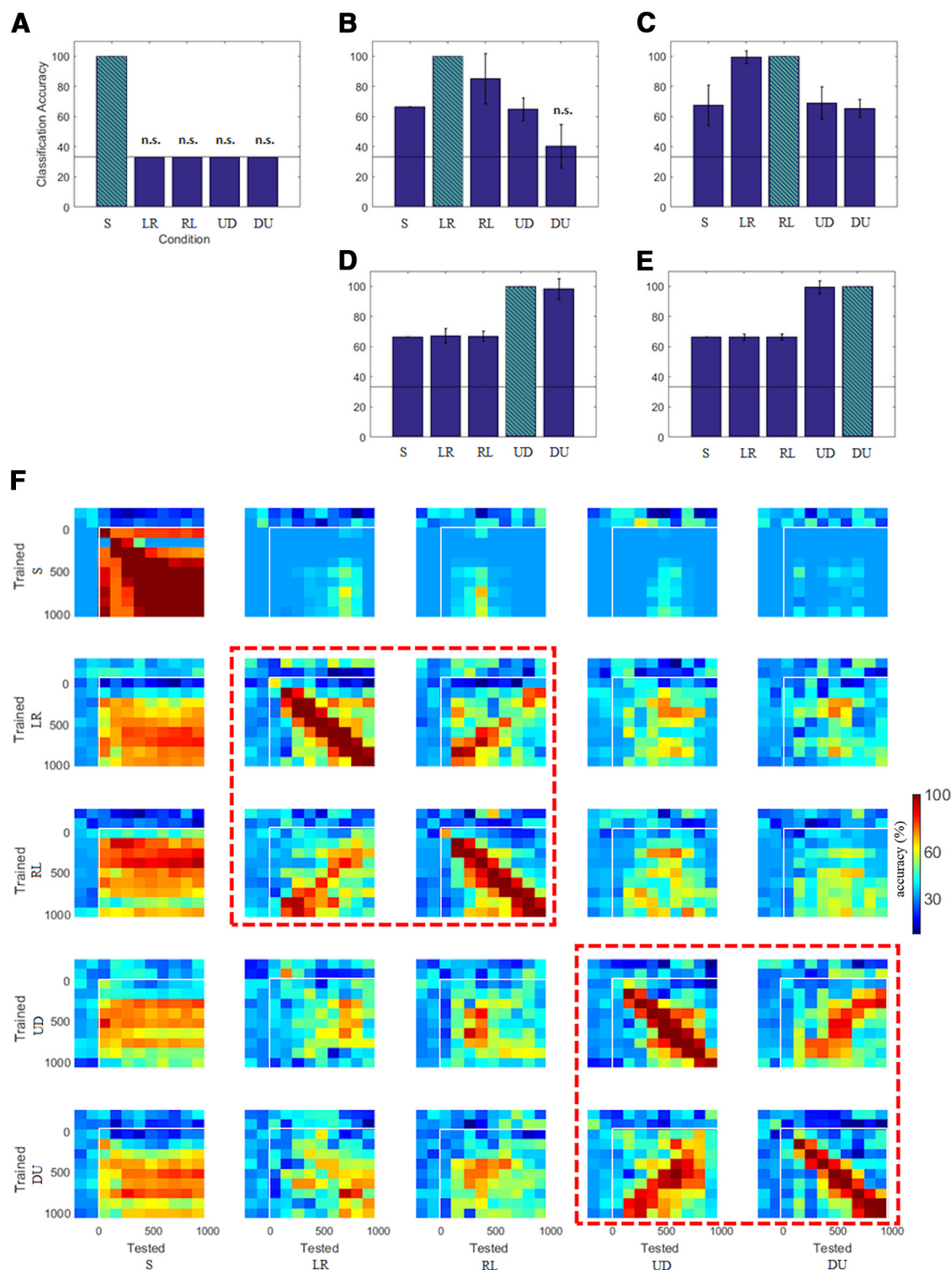


Figure 5. Decoding of shape identity from the responses of the recorded neurons ($n = 185$ neurons with 10 trials per condition). **A**, Classification accuracy (% correct) when the classifier was trained with the responses to static shapes using the 800-ms-long analysis window. Testing was performed for the static (S) and the four slit-viewing motion directions (LR, RL, UD, DU). Horizontal stippled line indicates chance performance (33%). **B**, Classification accuracies when training shape identification using the responses of the LR slit-viewing condition and testing each of the five conditions. Error bars indicate SD of the classification scores across resamplings. All classification scores were significant (permutation test with shuffled condition labels), except those indicated by n.s. **C**, Classification accuracies when training shape identification using the responses of the RL condition. Same conventions as in **A** and **B**. **D**, Classification accuracies when training shape identification using the responses of the UD condition. Same conventions as in **A** and **B**. **E**, Classification accuracies when training shape identification using the responses of the DU condition. Same conventions as in **A** and **B**. **F**, Classification scores when training and testing were done using bins of 100 ms duration. Training and testing were performed for all combinations of bins and conditions. Each panel represents the classification scores with the trained bin along the vertical axis and the tested bin along the horizontal axis. Panels of a row have the same trained condition, while the columns correspond to the tested conditions. White vertical and horizontal line in each panel indicates the onset of the first visible shape fragment. Red stippled square outlines indicate panels where trained and tested conditions had the same slit orientation. Percent correct classifications are presented with a color code. The same population of neurons as in **A–E**.

axes, which results in more diffuse, less organized cross-temporal decoding plots for the cross-axis classifications. Also, the overall classification accuracy for cross-axis generalization will be less than for same axis motion direction generalization since effective shape fragments may not be present for both axes in some single neurons. This is likely because the shapes were asymmetric. Temporal a-specific decoding was present when the slit-view conditions served as training data and the responses to the static whole-shape presentations were tested. This is expected since the effective shape fragments are present during the entire duration of the static presentation of the shape. There was only a weak generalization from static presentation to slit-views (Fig. 5F, top row panels), which agrees with the generalization data for the 800-ms-long analysis window (Fig. 5A).

Responses to shape fragments during slit-viewing and static presentation

The cross-temporal decoding analysis (Fig. 5F) suggests that the responses during slit-viewing are driven by effective shape fragments that become briefly visible. This would imply that the neurons respond to the thin shape strip that is visible through the narrow slit. We tested this directly by presenting 11 snapshots of the slit-viewing movie to a sample of neurons (in 2 monkeys) that were also tested during slit-viewing. The snapshots were presented briefly for 80 ms, each preceded and succeeded by the background noise pattern with the empty slit (see Materials and Methods). An example neuron tested in this snapshot test is shown in Figure 6B. It responded selectively to shape features related to the arms of an ape silhouette. To relate the responses in the snapshot test to those obtained during slit-viewing, we used a method akin to reverse correlation to compute the responses to individual shape fragments during slit-viewing (see Materials and Methods). This procedure yields a response for each frame of the slit-viewing movie, which can be visualized as a shape response plot in which the response to an individual shape fragment is indicated by a color code. This is illustrated for the same neuron in Figure 6A for both horizontal motion directions. This neuron responded for two brief periods during slit-viewing for both motion directions. Considering the typical response latency of IT neurons, the neuron responded during slit-viewing to parts of the arm of the ape silhouette. As shown in Figure 6B, the effective shape fragments obtained using reverse correlation of the slit-view data of each motion direction corresponded to those revealed by the snapshot test. When comparing the snapshot responses with the reverse-correlated slit-viewing responses, we reversed the plotted order of the fragments of one of the motion directions, so that shape fragments corresponded in the shape response plots of the two motion directions. Other examples of shape response plots for both motion directions and the corresponding snapshot test plots are shown in Figure 6D.

To quantify the correspondence between the responses during slit-viewing and in the snapshot test, we binned the responses of the shape response plots for the slit-viewing conditions in 11 bins and computed for each motion direction the Pearson correlation coefficient between the thus obtained responses for the slit-viewing and those of the snapshot test. For the neuron of Figure 6B, the correlation coefficients were close to 1, demonstrating the excellent fit between the responses to shape fragments during slit-viewing and in the snapshot test. We computed this correlation for all the snapshot test: slit-viewing combinations for which there was a significant response during slit-viewing (significance tested with ANOVA). A total of 45 and 15 neurons were tested in

Monkey MG and Monkey MT, respectively; and in 33 neurons, we had snapshot tests for both motion axes. Figure 6C shows the distribution of the correlation coefficients for each monkey separately, which was shifted toward positive values. The median correlation coefficients were significantly >0 in each monkey (Wilcoxon test; MG: median = 0.48; $p = 9 \times 10^{-27}$; MT: median = 0.40; $p = 1.6 \times 10^{-13}$), indicating that responses during slit-viewing and the static snapshot test are related. Thus, the observation that the single units responded during only brief periods of the slit-viewing can be explained by selective responses to effective shape fragments that were revealed by the slit during these moments.

Behavioral assessment of shape discrimination during slit-viewing: DMS task

The single-unit IT data reported above showed no evidence of temporal integration of the whole shape during slit-viewing. This raises the question of whether macaques perceive a shape during slit-viewing, as humans do. To examine this, we trained Monkey MG after the above-reported recordings in a DMS task in which he had to match a shape, presented during slit-viewing, and a static presentation of the same shape (Fig. 1D). Trials in which the sample stimulus was a whole shape were interleaved with slit-viewing samples. After training with a pool of 60 shapes, the monkey was tested with various combinations of the match and nonmatch stimuli, including novel shapes that were not used during the training period. Figure 7A summarizes the performance of the monkey for combinations of novel and old stimuli, presented either as sample or test stimuli. The old and novel stimulus sets consisted of 10 shapes each, and stimuli were randomly interleaved across trials. The number of sample presentations of a particular shape varied between 8 and 11. Although the DMS performance was greater when the sample stimulus was a static whole shape (82% correct; 95% CI = 77–88), the behavioral performance (70% correct; CI = 67–73) when the sample shape was presented during slit-viewing was well above chance (50% correct). There was no evidence of a difference in the performance between the novel and familiar, old shapes. Performance was highly similar for vertical (mean = 69%; CI = 62–75) and horizontal slit orientations (mean = 70%; CI = 63–76). Analyzing the performance for first trial presentations of a shape during slit-viewing also showed above-chance performance (67% correct; CI = 61–72), demonstrating that the above-chance performance for the slit-viewing sample conditions did not result from paired-associate learning of shape fragment samples and whole-shape match stimuli.

One could argue that the monkey used isolated shape fragments but no integrated whole-shape percept to match the slit-views and whole shapes. To examine this possibility, we presented trials in which the shape fragments were presented either in their correct order or in random order, as in our random slit-viewing conditions in the single-unit recording experiment. The original whole shape and a whole “random” shape, which was a spatial concatenation of the shape parts following the order of the random slit-views (as in Fig. 1C), served as the match and nonmatch stimuli. We reasoned that, if isolated shape features were driving the performance of the monkey, the monkey should show poor performance when having to choose between these match and nonmatch stimuli since both contained the same shape fragments as the sample stimulus. When the sample stimulus was an original slit-view, the performance of the monkey was 67% correct (CI = 63–70), which was well above chance level. This suggests that the monkey is not merely relying on isolated

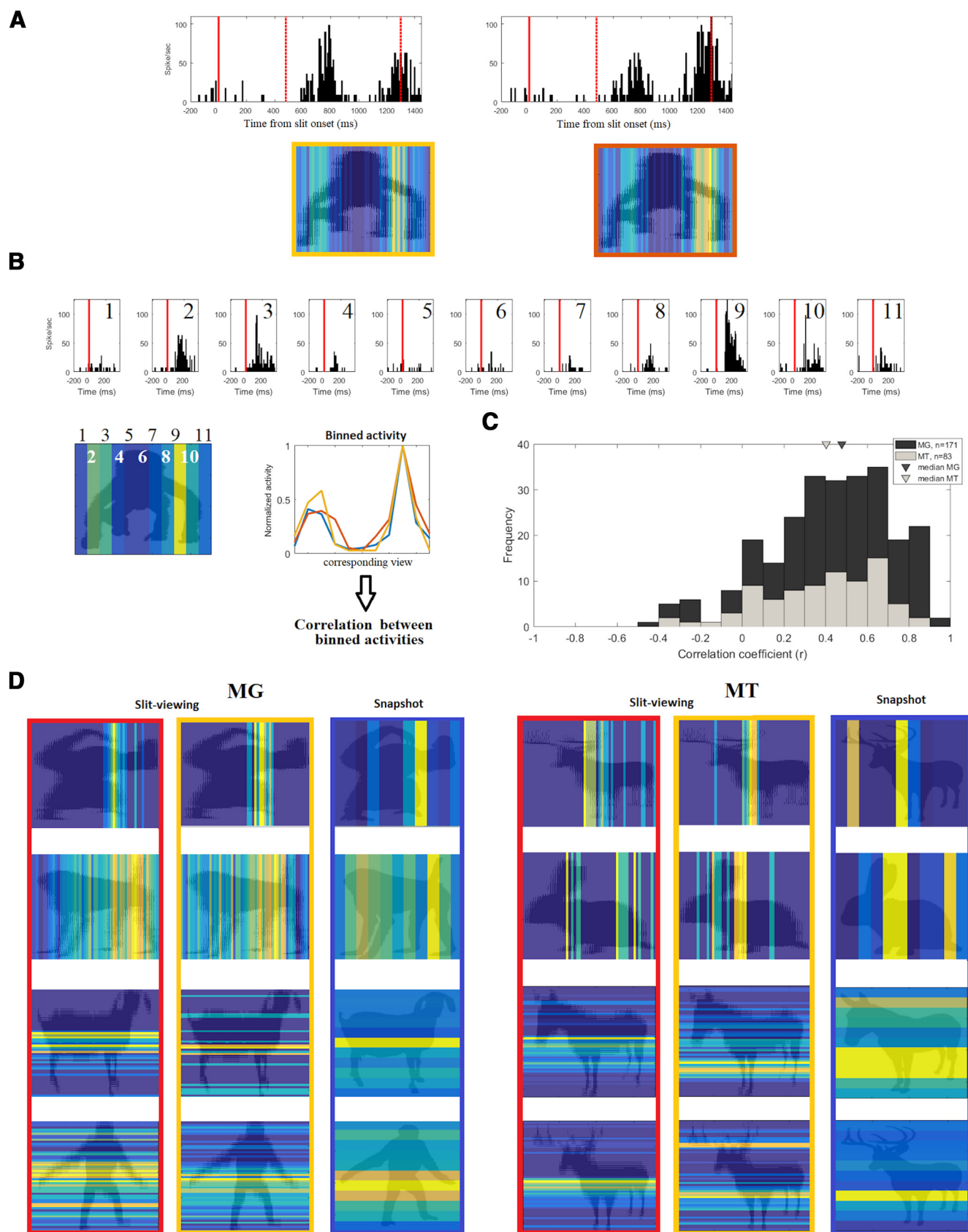


Figure 6. Responses in slit-viewing and snapshot tests compared. **A**, Example neuron, recorded in Monkey MG. PSTHs (bin width = 10 ms) show the responses of the neuron in the LR (left PSTH) and RL (right) original slit-viewing conditions. Conventions as in Figure 2. The panels below the PSTHs represent the responses to the shape segments of individual frames using a color code, with hot and cold colors representing a strong and weak response, respectively. The responses in these “shape response plots” were computed using a bin of 75 ms duration and a delay of 70 ms relative to shape fragment onset. **B**, The first row represents the responses in the snapshot test for the same neuron as in **A**. Full red line indicates stimulus onset. The shape response plot (second row) shows the normalized firing rate to the 11 shape fragments in color code, using the same convention as in **A**. Middle lower panel (binned activity) represents the mean

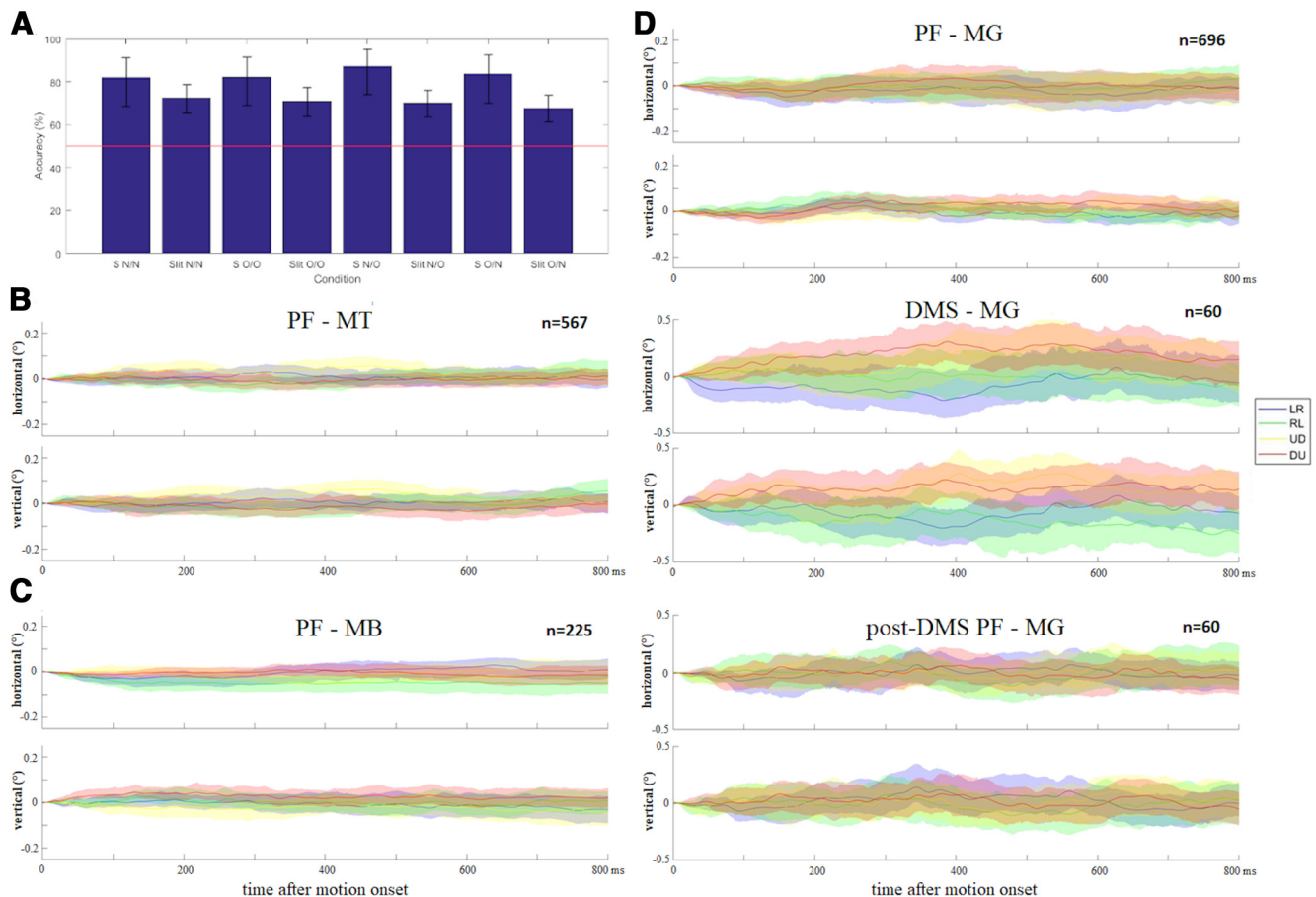


Figure 7. Behavioral results in the DMS task and eye movements. **A**, Behavioral accuracy (% correct) in the DMS task for different combinations of sample and test stimulus conditions. S, Static whole-shape sample presentation; Slit, slit-view sample stimulus; N, new stimulus; O, old stimulus; X, match stimulus; Y, nonmatch stimulus. Thus, the condition “Slit N/O” corresponds to trials in which the sample stimulus was a slit-view presentation of a new stimulus, the match was a new shape, and the nonmatch an old shape. Error bars indicate 95% CIs. Horizontal red line indicates the chance level (50% correct). **B**, Mean horizontal and vertical eye position and 95% CIs during slit-viewing in the PF task for each of the four motion directions, obtained during the recordings in Monkey MT. Before averaging, we subtracted for each trial the mean eye position in a 20–0 ms period before motion onset from the eye positions after motion onset. For each test session of a neuron, the baseline-subtracted eye positions were first averaged across the trials in which the same shape was presented and then averaged across condition–neuron combinations. The number of observations (n) corresponds to the number of condition–neuron combinations. **C**, Same as in **B**, but data of Monkey MB in the PF task. **D**, Eye position data for Monkey MG obtained during the recordings in the PF task before DMS task training (top panels), in the DMS task (middle panels), and the PF task after training of the DMS task (post-DMS; bottom panels). Same conventions as in **B**.

shape fragments when matching slit-views and test stimuli. When the sample stimulus was a random slit-viewing stimulus, the performance of the monkey dropped to 46% (CI = 43–50), which was statistically not different from chance. However, when the random stimulus configuration was shown as a static shape and the match and nonmatch stimuli were original and random

whole-shape configurations, the performance was 74% (CI = 67–80), which was highly similar to the performance when the original, whole shape was the sample stimulus (73%; CI = 66–79). Thus, although the monkey was able to match original and random whole-shape configurations, he was unable to match random-slit views. Such poor performance was also present when the random slit-views served as sample stimuli and both match and nonmatch stimuli were random shape configurations (mean = 54% correct; CI = 49–59). We attribute the chance performance for random slit-view samples to the difficulty of temporally integrating the randomly ordered shape parts, because of the large spatiotemporal discontinuities between successive fragments, into a single shape percept. These behavioral data support the presence of anorthoscopic shape perception in our slit-viewing conditions in macaques.

Single-unit responses during slit-viewing in DMS and PF tasks

To examine the possibility that temporal shape integration during slit-viewing at the IT single-unit level would occur when the monkey is attending the stimulus, we recorded 20 responsive neurons while Monkey MG was performing the DMS task using

normalized firing rate for 11 bins of the shape response plots for each of the two motion directions presented in **A** and the snapshot condition. We reversed the order of the shape segments of the RL slit-viewing condition (yellow) to have corresponding shape parts in the different vertical slit conditions (LR slit-viewing [red] and snapshot [blue] test conditions). The responses were normalized by the maximum response for each condition separately, and Pearson correlation coefficient (r) was calculated between the binned responses of the slit-viewing conditions and the snapshot condition. **C**, Distribution of the Pearson correlation coefficient between the binned responses to the shape fragments during slit-viewing and in the snapshot test. Data of the 2 monkeys are shown separately. Colored triangles represent the medians. The number of observations (n) corresponds to slit-viewing direction–shape combinations. **D**, Shape response plots for two opponent motion directions of the slit-viewing and corresponding snapshot tests for other example neurons (rows) recorded in Monkeys MG and MT. The same conventions as in **B**. Examples of vertical (top two rows) and horizontal (bottom two rows) slit orientations are shown for each monkey.

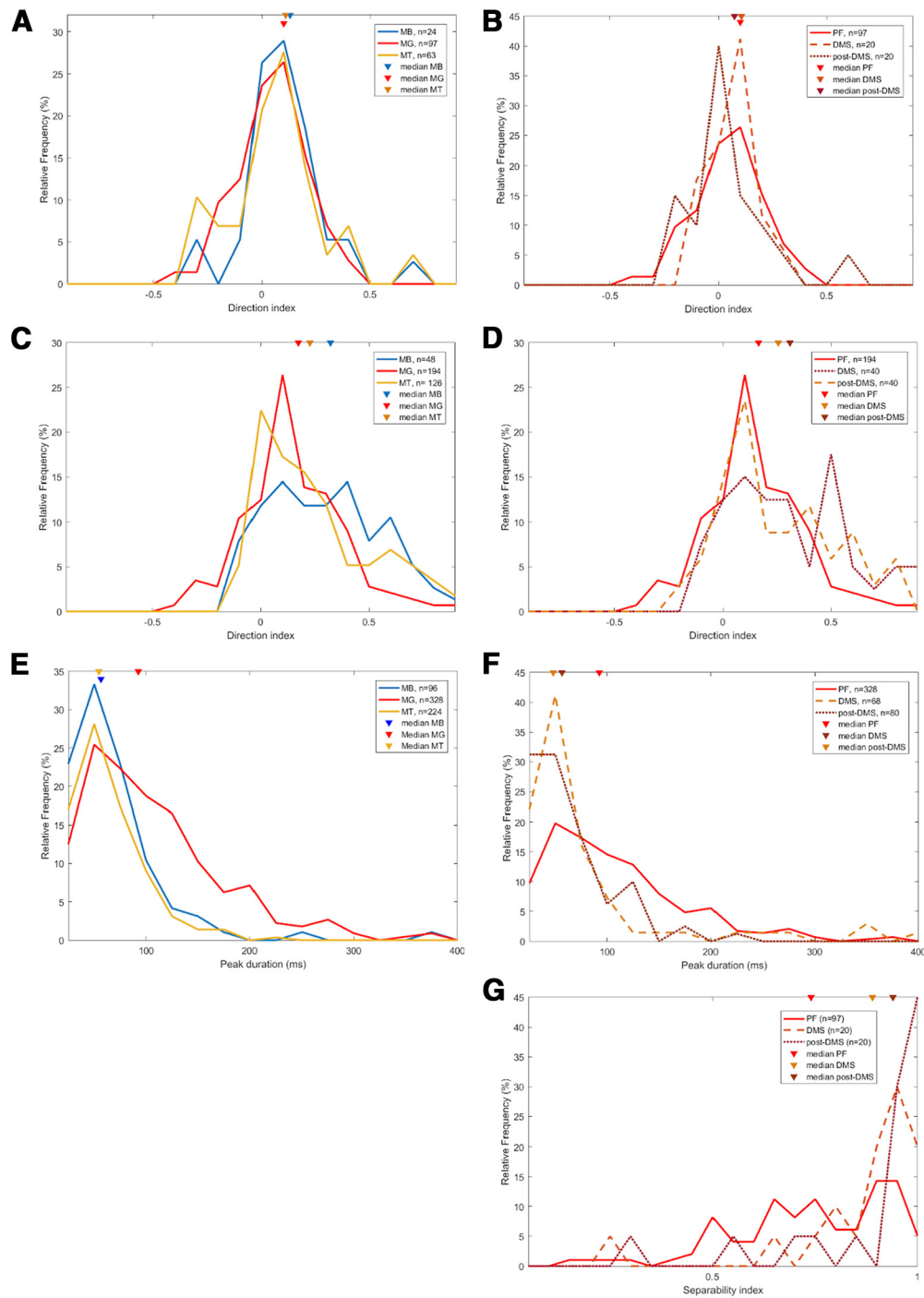


Figure 8. Distributions of response properties in the PF and DMS tasks during slit-viewing. **A**, Distribution of the same axis Direction Index for the neurons recorded in the PF task for each of the 3 monkeys. Triangles represent medians. **B**, Distribution of the same axis Direction Index for neurons recorded in the PF task before training, during the DMS task, and in the post-DMS task. Data of Monkey MG. **C**, Distribution of the Direction Index for orthogonal axes for the neurons recorded in the PF task for each of the 3 monkeys. Each neuron contributed two Direction Index values, one for each of the two directions along the orthogonal axis. **D**, Distribution of the Direction Index for orthogonal axes for neurons recorded in the PF task before training, during the DMS task, and in the post-DMS task. Data of Monkey MG. **E**, Distribution of the peak duration metric for the neurons recorded in the PF task for each of the 3 monkeys. **F**, Distribution of the peak duration metric for the neurons recorded in the PF task before DMS training, during the DMS task, and post-DMS training in Monkey MG. **G**, Distribution of the Separability Index for the neurons recorded in the PF task before DMS training, during the DMS task, and post-DMS training in Monkey MG.

slit-views and static whole shapes as sample stimuli. The sample stimuli in the DMS task were three shapes that were selected anew for each neuron based on the responses in the search test in which 70 shapes were presented during PF. After the recordings during the DMS task, we also recorded an additional 20 responsive neurons in Monkey MG while he was performing the same slit-viewing test with PF as before the DMS training. To

compare quantitatively the three samples of neurons, those recorded during PF, while performing the DMS task, and during post-DMS PF, we computed five response property indices for each neuron using data for the original shape slit-viewing conditions. The first three indices quantified the effect of motion direction on the responses during slit-viewing of the best shape. The first of these Direction Indices compared the response to the

best motion direction and its opponent direction for the same motion axis (see Materials and Methods). For each neuron, its best direction was determined in independent trials, explaining the presence of negative Direction Indices. For the three samples of neurons of MG, the median Direction Indices for same axis directions were low (Fig. 8B) and similar to those obtained in the other 2 monkeys during PF (Fig. 8A). The low median values (~ 0.1 , or a 22% difference) agree with the finding that the responses during slit-viewing are driven by effective shape fragments, which are identical for the two directions of the same axis. However, for all five samples, the same axis Direction Index was positive, and significantly greater than zero in each monkey (MG: $p = 1.2 \times 10^{-5}$; MT: $p = 3.7 \times 10^{-3}$; MB: $p = 5.8 \times 10^{-5}$; Wilcoxon test), showing an albeit weak influence of motion direction on the responses. This agrees with previous studies that showed stimulation history effects on IT responses (see Discussion).

The other Direction Indices compared the best direction and the two directions along the orthogonal motion axis. For each neuron, we computed two such indices, one for each orthogonal motion direction, and these were pooled in the analyses. Since effective shape fragments for the two slit orientations can drive a neuron to a different extent, we expected the Direction Indices for different axes to be greater than for Direction Indices of the same motion axes, which was indeed the case for each of the 3 monkeys (Fig. 8C). Importantly, this also was the case after training and during the DMS task in Monkey MG (Fig. 8D). Indeed, the median Direction Index for orthogonal axes was significantly smaller in the PF task before than after the DMS training ($p = 0.0017$; Wilcoxon rank sum test; data of DMS and post-DMS tasks pooled), which is opposite to what one would expect when training or DMS task execution would have improved temporal shape integration that generalized across slit orientation.

We also quantified for each neuron the period during which the neuron responded during slit-viewing by measuring the duration of its response at half-height (see Materials and Methods). This peak duration estimation was performed for the slit-viewing condition that produced the maximum response of the neuron. One would predict longer peak durations after DMS training or during DMS task performance when this increased temporal shape integration. However, the opposite trend was present (Fig. 8E; PF vs DMS: $p = 0.0035$; PF vs post-DMS: $p = 0.0018$; Wilcoxon rank sum test), suggesting that temporal shape integration did not increase during or after the DMS task. The median peak durations during the DMS task and post-DMS in Monkey MG are similar to those obtained during the PF task in the other 2 monkeys (Fig. 8E).

To assess the extent to which shape and motion direction were encoded in a separable manner, we computed for each neuron the Separability Index (see Materials and Methods). A high Separability Index indicates that one can predict the responses in a slit-viewing condition by knowing the responses to the shape, irrespectively of motion direction, and to the motion direction, irrespectively of shape. High separability will ensure invariant decoding of the shape, irrespectively of motion direction (Li et al., 2009). The Separability Indices were computed after integrating the responses during slit-viewing using the 800 ms analysis window. We found that the median Separability Index was significantly greater when the monkey was performing the DMS task compared with the PF before the training ($p = 5.3 \times 10^{-3}$; Wilcoxon rank sum test; Fig. 8G). This was not because of the difference between the two tasks, since the neurons recorded post-training during PF showed also a greater

median Separability Index than the sample of neurons recorded before the DMS training ($p = 7.9 \times 10^{-4}$; Wilcoxon rank sum test; Fig. 8G). One possible explanation for the increased Separability Index after DMS training is that the shape selectivity of the neurons was greater post-training. To address this possibility, we computed for each neuron a Shape Selectivity Index, being the response to the best static whole shape minus the response to the worst static whole shape, divided by the sum of the responses to both shapes. The median Shape Selectivity Index was indeed significantly greater after (median index during DMS task: 0.71; post-DMS: 0.9) compared with before DMS training (median: 0.57; $p = 1.2 \times 10^{-6}$; Wilcoxon rank sum test; data of DMS and post-DMS tasks pooled). Thus, the shape selectivity was higher for the sample of neurons recorded after the DMS training, which can explain their higher Separability Indices.

Eye movements

Analysis of the eye movements during slit-viewing showed a rather stable fixation for the different motion directions in each monkey while performing the PF task (Fig. 7B–D). Monkey MG showed somewhat higher variability in eye positions during the slit-view presentations of the DMS task (Fig. 7D, middle), but there was no evidence of smooth pursuit of the motion of the shape.

Discussion

We recorded the responses of body patch (ASB) neurons to moving shapes that were only partially visible through a static narrow slit. Although only a small fragment of the shape was revealed through the slit at a single moment in time, the population of IT neurons signaled shape identity by their response when that was cumulated across the viewing period. The shape preference for the slit-viewing conditions was the same as for static whole-shape presentations. However, by analyzing the responses on a finer time scale and comparing responses between motion directions, we showed that IT neurons responded to particular shape features that were visible through the slit. Furthermore, the responses during slit-viewing were predicted by the responses of the same neuron to static presentations of shape fragments as revealed through the slit. In this IT body patch, we found no evidence for temporal integration of the sequentially revealed shape parts into a coherent shape representation, neither at the single unit nor the population response level. Qualitatively identical response patterns were present when a monkey was matching slit-views of a shape to static whole-shape presentations and thus attending the slit-views. These data suggest that, although the temporally integrated response of macaque IT neurons can signal shape identity under slit-viewing, the temporal integration needed for the formation of a whole-shape percept occurs in other areas, perhaps downstream to IT.

The following observations led to the conclusion that single IT neurons responded to partial shape views but did not integrate the whole shape, during slit-viewing. First, single IT neurons responded only during particular periods of the slit-viewing movie, and these corresponded to periods in which effective shape fragments were displayed as determined in an independent test with static snapshots of the movie. Second, shape decoding during slit-viewing was restricted in time and generalized across opposite motion directions for the different periods in which the same shape fragment was presented for the two directions. Third, the generalization of shape decoding was less across axes

than across directions of the same axis, which is expected when the neurons respond to shape fragments because shape features overlap less for different slit orientations. However, whole-shape representations are expected to be invariant to slit orientation. Furthermore, shape selectivity did not consistently increase during slit-viewing. Fourth, the average shape preference was the same for the random and original slit-viewing conditions, supporting the conclusion that the neurons represent shape fragments.

The response to the slit-views could differ, albeit weakly, between opposite motion directions, although shape fragments were equal for the two directions. This is not surprising since the response to a particular shape fragment will also be determined by the response selectivity of the neuron for preceding shape features, which will differ between directions. Indeed, responses of IT neurons depend on stimulation history and preceding stimuli, for example, adaptation (Vogels, 2016) and shape sequence effects as observed for forward versus backward walking sequences (Vangeneugden et al., 2011). Although these effects of preceding stimulation history and the motion direction sensitivity found in the present study show that IT neurons show temporal integration of preceding and current stimulation, this should be distinguished from temporal integration of dynamic partial shape views for the formation of a whole-shape percept during slit-viewing. The latter requires temporal integration and maintenance of shape information across slit-views, taking into account the relative location of the shape fragments based on the velocity of the features in the slit during stimulation (Ögmen and Herzog, 2016). Our results are in line with previous estimates of a temporal integration duration of 100–120 ms in the macaque rostral STS for visual action sequences (Singer and Sheinberg, 2010; Vangeneugden et al., 2011), which is shorter than required for integration of slit-views into a whole-shape percept.

Although the recorded neurons did not temporally integrate the slit-views, a linear classifier could classify the shapes when it had as input the responses averaged over the whole slit-viewing period. Furthermore, for this averaged response, a generalization of shape classification occurred across motion directions, across motion axes, and for static whole-shape and slit-view presentations. These results agree with a human fMRI study in which generalization of shape classification across slit-viewing axes and for static whole shape was shown with multivoxel pattern analysis in ventral stream visual areas (Orlov and Zohary, 2018). Because of the temporally coarse hemodynamic response, the BOLD response amounts to using a long window in which responses will be temporally integrated, similar to what we did when averaging the responses over the slit-viewing period. However, our results show that such generalization of classification using the integrated response cannot be used as evidence for a temporally integrated whole-shape representation at the level of single neurons. In contrast to the presence of shape selectivity in the random slit-viewing conditions in our macaque data, classification of shape based on random slit-view presentations was at chance level in the human fMRI study (Orlov and Zohary, 2018). However, they randomized the frames at a 60 Hz rate in their random condition. This amounts to a rapid serial presentation of randomly ordered shape fragments at 60 Hz. In macaque STS, such rapid serial presentation decreases stimulus-selective responses (Keysers et al., 2005; De Baene et al., 2007) because of forward and backward masking, and we expect the same reduction in human ventral stream areas. This may explain why shape classification with multivoxel pattern analysis of LOC activations was at chance level in the random slit-viewing conditions (Orlov and Zohary, 2018). As noted above, moving shapes can produce

stronger responses because of the similarity of the nearby stimulus features during slit-viewing.

One behavioral study suggested poor spatiotemporal integration during slit-viewing in apes compared with humans (Imura and Tomonaga, 2013), but such quantitative interspecies comparisons are difficult to interpret since nonperceptual, cognitive differences between species can affect visual task performance. What that study did show is that chimpanzees perform better than the chance level when matching dynamic slit-views to static whole-shape outlines. We found a similar result here in a macaque, suggesting that nonhuman primates show AP. This also suggests that the lack of a temporally integrated whole-shape representation at the level of macaque IT neurons during slit-viewing is not because monkeys do not show AP. The single-unit properties were similar when the monkey was performing matching of slit-views and thus attending the slit-view presentations. This observation, together with the presence of fMRI activations in LOC during the performance of an orthogonal task during slit-viewing (Yin et al., 2002), suggests that the use of a PF task in most of our recordings cannot explain the lack of temporal whole-shape integration for the IT responses.

In addition, there is no reason to assume that our findings depended on the choice of recording from the body patch ASB since Orlov and Zohary (2018) reported whole-shape representations during slit-viewing in the large expanse of LOC and even face/body-selective regions, and AP is present for animal shapes (Parks, 1965). IT neurons remain shape-selective under conditions in which a static or moving pattern occludes partially a shape (Kovacs et al., 1995b). Multiple fragments of the shape are presented simultaneously in such occlusion displays; and thus, no temporal integration is required to obtain shape completion. Selective responses to face or body parts were reported in other studies posterior to ASB when only small fragments of a stimulus were presented (Bubbles) to assess the feature selectivity of posterior face (Issa and Dicarlo, 2012) and body patch neurons (Popivanov et al., 2016). Unlike in the present study, no whole-face or body percept is present in such reduced displays.

Although our findings suggest that single IT neurons do not form whole-shape representations during slit-viewing, their responses, integrated across slit-viewing, contain sufficient information for shape identification. This begs the question of where such temporal integration occurs and even whether IT plays a role in building a whole-shape representation during slit-viewing. Psychophysical studies show that AP depends on the estimation of shape velocity (Morgan et al., 1982; Shimono and Richards, 1986), which may imply the contribution of dorsal visual areas in the formation of a shape percept during slit-viewing. Indeed, human fMRI studies showed dorsal visual area (e.g., hMT/V5) activations during slit-viewing, although these were less than in ventral areas (Yin et al., 2002; Orlov and Zohary, 2018). An area that underlies spatiotemporal shape integration during slit-viewing needs to be able to maintain and update shape information during viewing in “nonretinotopic” memory (Ögmen and Herzog, 2016), which requires temporal integration with a long time constant. Because temporal integration constants increase along the cortical hierarchy (Hasson et al., 2008; Murray et al., 2014; Spitmaam et al., 2020), it is possible that higher-order regions, such as the PFC or more anterior in temporal cortex than ASB, might underlie whole-shape formation during slit-viewing, but testing this requires further work. Finally, we note that current computational models of the ventral visual stream (Kar et al., 2019) can accommodate the observed responses in macaque IT during slit-viewing. However, to explain AP computational models of visual recognition needs to be augmented.

References

- Bao P, She L, McGill M, Tsao DY (2020) A map of object space in primate inferotemporal cortex. *Nature* 583:103–108.
- De Baene W, Premereur E, Vogels R (2007) Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. *J Neurophysiol* 97:2900–2916.
- Denys K, Vanduffel W, Fize D, Nelissen K, Peuskens H, Van Essen D, Orban GA (2004) The processing of visual shape in the cerebral cortex of human and nonhuman primates: a functional magnetic resonance imaging study. *J Neurosci* 24:2551–2565.
- Ekstrom LB, Roelfsema PR, Arsenault JT, Bonmassar G, Vanduffel W (2008) Bottom-up dependent gating of frontal signals in early visual cortex. *Science* 321:414–417.
- Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N (2008) A hierarchy of temporal receptive windows in human cortex. *J Neurosci* 28:2539–2550.
- Imura T, Tomonaga M (2013) Differences between chimpanzees and humans in visual temporal integration. *Sci Rep* 3:3256.
- Issa EB, Dicarlo JJ (2012) Precedence of the eye region in neural processing of faces. *J Neurosci* 32:16666–16682.
- Kar K, Kubilius J, Schmidt K, Issa EB, DiCarlo JJ (2019) Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat Neurosci* 22:974–983.
- Keysers C, Xiao DK, Foldiak P, Perrett DI (2005) Out of sight but not out of mind: the neurophysiology of iconic memory in the superior temporal sulcus. *Cogn Neuropsychol* 22:316–332.
- Kovacs G, Vogels R, Orban GA (1995a) Cortical correlate of pattern backward masking. *Proc Natl Acad Sci USA* 92:5587–5591.
- Kovacs G, Vogels R, Orban GA (1995b) Selectivity of macaque inferior temporal neurons for partially occluded shapes. *J Neurosci* 15:1984–1997.
- Kriegeskorte N (2015) Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu Rev Vis Sci* 1:417–446.
- Kumar S, Popivanov ID, Vogels R (2019) Transformation of visual representations across ventral stream body-selective patches. *Cereb Cortex* 29:215–229.
- Li N, Cox DD, Zoccolan D, Dicarlo JJ (2009) What response properties do individual neurons need to underlie position and clutter 'invariant' object recognition? *J Neurophysiol* 102:360–376.
- McCloskey M, Watkins MJ (1978) The seeing-more-than-is-there phenomenon: implications for the locus of iconic storage. *J Exp Psychol Hum Percept Perform* 4:553–564.
- Meyers EM (2013) The neural decoding toolbox. *Front Neuroinform* 7:8.
- Morgan MJ, Findlay JM, Watt RJ (1982) Aperture viewing: a review and a synthesis. *Q J Exp Psychol A* 34:211–233.
- Murray JD, Bernacchia A, Freedman DJ, Romo R, Wallis JD, Cai X, Padoa-Schioppa C, Pasternak T, Seo H, Lee D, Wang XJ (2014) A hierarchy of intrinsic timescales across primate cortex. *Nat Neurosci* 17:1661–1663.
- Mysore SG, Vogels R, Raiguel SE, Todd JT, Orban GA (2010) The selectivity of neurons in the macaque fundus of the superior temporal area for three-dimensional structure from motion. *J Neurosci* 30:15491–15508.
- Ögmen H, Herzog MH (2016) A new conceptualization of human visual sensory-memory. *Front Psychol* 7:830.
- Orlov T, Zohary E (2018) Object representations in human visual cortex formed through temporal integration of dynamic partial shape views. *J Neurosci* 38:659–678.
- Parks TE (1965) Post-retinal visual storage. *Am J Psychol* 78:145–147.
- Popivanov ID, Jastorff J, Vanduffel W, Vogels R (2012) Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI. *Neuroimage* 63:723–741.
- Popivanov ID, Jastorff J, Vanduffel W, Vogels R (2014) Heterogeneous single-unit selectivity in an fMRI-defined body-selective patch. *J Neurosci* 34:95–111.
- Popivanov ID, Schyns PG, Vogels R (2016) Stimulus features coded by single neurons of a macaque body category selective patch. *Proc Natl Acad Sci USA* 113:E2450–E2459.
- 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.
- Rock I, Sigman E (1973) Intelligence factors in the perception of form through a moving slit. *Perception* 2:357–369.
- Sawamura H, Orban GA, Vogels R (2006) Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the FMRI adaptation paradigm. *Neuron* 49:307–318.
- Shimojo S, Richards W (1986) 'Seeing' shapes that are almost totally occluded: a new look at Parks's camel. *Percept Psychophys* 39:418–426.
- Singer JM, Sheinberg DL (2010) Temporal cortex neurons encode articulated actions as slow sequences of integrated poses. *J Neurosci* 30:3133–3145.
- Spitmaam M, Seo H, Lee D, Soltani A (2020) Multiple timescales of neural dynamics and integration of task-relevant signals across cortex. *Proc Natl Acad Sci USA* 117:22522–22531.
- Taubert J, Van Belle G, Vanduffel W, Rossion B, Vogels R (2015) Neural correlate of the Thatcher face illusion in a monkey face-selective patch. *J Neurosci* 35:9872–9878.
- Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. *Nat Neurosci* 6:989–995.
- Van den Hurk J, Op de Beeck HP (2019) Generalization asymmetry in multivariate cross-classification: when representation A generalizes better to representation B than B to A. *bioRxiv* 592410.
- Vanduffel W, Fize D, Mandeville JB, Nelissen K, Van Hecke P, Rosen BR, Tootell RB, Orban GA (2001) Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. *Neuron* 32:565–577.
- Vangeneugden J, De Maziere PA, Van Hulle MM, Jaeggli T, Van GL, Vogels R (2011) Distinct mechanisms for coding of visual actions in macaque temporal cortex. *J Neurosci* 31:385–401.
- Vinken K, Op de Beeck HP, Vogels R (2018) Face repetition probability does not affect repetition suppression in macaque inferotemporal cortex. *J Neurosci* 38:7492–7504.
- Vogels R (2016) Sources of adaptation of inferior temporal cortical responses. *Cortex* 80:185–195.
- Yamins DL, DiCarlo JJ (2016) Using goal-driven deep learning models to understand sensory cortex. *Nat Neurosci* 19:356–365.
- Yin C, Shimojo S, Moore C, Engel SA (2002) Dynamic shape integration in extrastriate cortex. *Curr Biol* 12:1379–1385.