

*Research Articles: Behavioral/Cognitive*

## Moving a Shape behind a Slit: partial Shape Representations in Inferior Temporal Cortex

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

**Cite as:** J. Neurosci 2021; 10.1523/JNEUROSCI.0348-21.2021

Received: 15 February 2021

Revised: 4 June 2021

Accepted: 8 June 2021

---

*This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.*

**Alerts:** Sign up at [www.jneurosci.org/alerts](http://www.jneurosci.org/alerts) to receive customized email alerts when the fully formatted version of this article is published.

1   **Moving a shape behind a slit: partial shape representations in inferior temporal cortex.**

2

3   Anna Bognár<sup>1,2</sup>, Rufin Vogels<sup>1,2</sup>

4

5       <sup>1</sup>. Laboratorium voor Neuro- en Psychofysiologie, Department of Neurosciences, KU  
6       Leuven, 3000 Leuven, Belgium

7       <sup>2</sup>. Leuven Brain Institute, KU Leuven, 3000 Leuven, Belgium.

8

9   Corresponding author: Rufin.vogels@kuleuven.be

10

11   Abbreviated title: inferior temporal responses during slit-viewing.

12

13   Number of pages: 48

14   Number of figures: 8

15   Number of words for abstract: 249

16   Number of words for introduction: 649

17   Number of words for discussion: 1497

18

19

20   Conflict of interest: The authors declare no competing financial interests.

21

22

23   Acknowledgments: The authors thank P. Kayenbergh, I. Puttemans, A. Hermans, G.  
24   Meulemans, W. Depuydt, C. Ulens, S. Verstraeten, and M. Depaepe for technical and  
25   administrative support. This research was supported by Fonds voor Wetenschappelijk  
26   Onderzoek (FWO) Vlaanderen (Odysseus G0007.12), KU Leuven grant C14/17/109 and  
27   received funding from the European Research Council (ERC) under the European Union's  
28   Horizon 2020 research and innovation programme (grant agreement No. 856495).

29   **Abstract**

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

50

51

52

53

54

55 **Significance Statement**

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

66

67

68

69       **Introduction**

70       Current models of the ventral visual stream (Kriegeskorte, 2015; Yamins and DiCarlo,  
71       2016; Kar et al., 2019) are based on spatial representations of an object's image. For  
72       instance, the activation of units of convolutional neural network models depends on the  
73       spatial integration of local stimulus features. However, one can recognize an object when it  
74       moves behind a static occluder and only a small part of its shape is visible at a given moment  
75       in time i.e., seeing a dog walk behind a slightly open door (PARKS, 1965; Rock and Sigman,  
76       1973). Psychophysical studies showed that the perception of a complete moving figure when  
77       it only is revealed, one tiny part at a time, through a static slit (anorthoscopic perception  
78       (AP)) is not due to retinal smearing because of pursuit eye movements but reflect temporal  
79       integration of spatially fragmented shape information (McCloskey and Watkins, 1978;  
80       Morgan et al., 1982). AP provides a challenge to current spatial-based models of object  
81       recognition since successive shape elements stimulating the same retinal strip must be  
82       integrated over time to obtain a representation of the object.

83       Human fMRI studies reported activations in dorsal and ventral visual stream areas  
84       during AP (Yin et al., 2002; Reichert et al., 2014; Orlov and Zohary, 2018). In particular, these  
85       fMRI studies suggested that, amongst other areas (e.g. hMT/V5), the Lateral Occipital  
86       Complex (LOC) area, a key area of the human ventral visual stream, is more active when  
87       through a narrow slit a whole object is perceived than when it is seen as isolated shape  
88       fragments. One recent fMRI study (Orlov and Zohary, 2018) presented non-familiar shapes  
89       that moved behind either a vertically or horizontally oriented narrow slit. Using multi-voxel  
90       pattern analysis (MVPA), they found that the pattern of activation in LOC encoded the shape

91 of the object when it was moving behind the narrow slit. This suggested to the authors that  
92 LOC represents a whole-shape percept based on the temporal integration of the slit-views.

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

108 To examine the time course of the responses during AP, we recorded single neurons  
109 of the macaque inferior temporal (IT) cortex during slit-viewing of shapes. Macaque IT is  
110 assumed to be the homolog of human LOC (Denys et al., 2004). To increase the effectiveness  
111 of the shapes in driving selective activity, we performed recordings in and close to an fMRI-  
112 defined body patch of the anterior superior temporal sulcus (ASB; (Kumar et al., 2017),  
113 employing silhouettes of animals, known to produce strong selective responses in that patch

114 (Bao et al., 2020). The shapes were presented statically or when moving behind a static  
115 narrow vertical or horizontal slit. Analysis of single-unit and population responses showed  
116 that the IT neurons responded selectively to shape fragments during slit-viewing but did not  
117 temporally integrate the shape, even when the monkey was matching static whole shapes to  
118 shapes presented during slit-viewing.

119

120 **Materials and Methods**

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

127 *fMRI body patch localizer.* The monkeys were scanned on a 3T Siemens Trio scanner  
128 following published standard procedures (Vanduffel et al., 2001). Functional MR images  
129 were acquired using a custom-made 8-channel monkey coil (Ekstrom et al., 2008) and a  
130 gradient-echo single-shot echo planar imaging sequence (for more details, see (Popivanov et  
131 al., 2012)). In monkeys MB and MT, we used the block design procedure of (Popivanov et al.,  
132 2012), showing 20 images of monkey bodies, monkey faces, objects, mammals, birds, and  
133 fruits/vegetables while the subjects were performing a passive fixation task for a juice  
134 reward. ASB is the most anterior body patch in the Superior temporal sulcus, defined by the  
135 contrast monkey bodies minus objects (for further details, see (Kumar et al., 2017) ). The  
136 fMRI data of MG were obtained in the context of previous studies (Taubert et al., 2015;  
137 Vinken et al., 2018). This monkey was scanned during passive fixation with stimuli of

138 different classes that were identical to those used by (Tsao et al., 2003). The contrast to  
139 define ASB was bodies (without heads) minus faces, fruits, tools, and hands. The fMRI maps  
140 were co-registered with an anatomical MRI of each monkey and these images were  
141 employed to position recording chambers and guide tube locations. Further details about  
142 the procedure used to target fMRI-defined body patches can be found in our previous  
143 publications ((Popivanov et al., 2014); (Kumar et al., 2017)).

144         *Electrophysiological recordings.* Standard single-unit recordings were performed with  
145 epoxylite-insulated tungsten microelectrodes (FHC, Maine, USA; in situ measured impedance  
146 about  $1M\Omega$ ) using techniques as described previously (Sawamura et al., 2006). Briefly, the  
147 electrode was lowered with a Narishige microdrive into the brain using a stainless steel  
148 guide tube that was fixed in a standard Crist grid positioned within the recording chamber.  
149 After amplification and filtering between 540 Hz and 6 kHz, spikes of a single unit were  
150 isolated online using a custom amplitude- and time-based discriminator. The recording grid  
151 locations were defined so that the electrode targeted ASB or neighboring sites in the left  
152 hemisphere. We employed the body patch localizer to increase the frequency of finding  
153 neurons that responded strongly and selectively to the employed shapes. Previous studies  
154 showed that ASB neurons respond well to images of four-legged animals (Kumar et al., 2017;  
155 Bao et al., 2020) and ASB is strongly activated by silhouettes of animals (Bao et al., 2020).  
156 Thus, by using silhouettes of animals as shapes and by targeting ASB and surrounding sites as  
157 recording locations we aimed to increase the efficiency of the recordings.

158         The position of one eye was continuously tracked using an infrared video-based  
159 tracking system (SR Research EyeLink, USA; sampling rate 1 kHz). Stimuli were displayed on a  
160 CRT display (Philips Brilliance 202 P4; 1024 x 768 screen resolution; 75 Hz vertical refresh  
161 rate) at a distance of 57 cm from the monkey's eyes. The on- and offset of the stimuli was

162 signaled using a photodiode detecting luminance changes of a small square in the corner of  
163 the display (but invisible to the monkey), placed in the same frame as the stimulus events. A  
164 Digital Signal Processing -based computer system developed in-house controlled stimulus  
165 presentation, event timing, and juice delivery while sampling the photodiode signal, vertical  
166 and horizontal eye positions, spikes, and behavioral events. Timestamps of the recorded  
167 spikes, eye positions, bandpass filtered electrode signal (sampling rate 40 kHz), stimulus, and  
168 behavioral events were stored for offline analyses. Isolation of the single units was checked  
169 offline using the spike-sorting software of the Spike2 analysis package.

170         *Stimuli and tasks.*

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

179         Search test. The trial started with the onset of a small red square (size 0.2°) on top of  
180 a 15° sized square that consisted of visual noise, presented on the gray background of the  
181 display. The static noise was created by randomly positioning white and black pixels ("salt  
182 and pepper" noise; pixel size = 0.03°). The monkeys had to fixate the red target for 250 ms.  
183 Then, the shape was presented for 350 ms on the noise background, centered behind the  
184 fixation target. The monkey had to continue fixation during the stimulus presentation and  
185 for a period of 108 ms after stimulus offset. Continuous fixation within a fixation window of

186  $2^\circ \times 2^\circ$  was rewarded with a drop of juice. The 70 shapes were shown in a pseudo-random  
187 order for at least 5 unaborted trials each. All neurons were tested using this procedure and  
188 the neuron was examined in further tests when a response was notable in the online Peri-  
189 stimulus Time Histograms (PSTHs) for at least one of the shapes. Based on the responses to  
190 the individual shapes, we selected a shape that produced the highest response ("best"), a  
191 shape for which there was no or a weak response ("worst" shape), and, for most recordings,  
192 also a shape that produced a response intermediate ("medium") between the best and  
193 worst shape.

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

208 The shape fragments shown in the slit depended on the condition. In the slit-viewing  
209 with original shape conditions ("original" condition), one of the selected shapes was moving

210 smoothly behind the slit with a speed of  $6.2^{\circ}/s$ . The slit was oriented either vertical or  
211 horizontal and presented at an eccentricity of  $2^{\circ}$  in the contralateral visual field (centered on  
212 the horizontal meridian) or below the fixation target (centered on the vertical meridian),  
213 respectively. In the case of the vertically oriented slit, the shape was either moving left- or  
214 rightwards. In the case of the horizontally oriented slit, the shape was moving up- or  
215 downwards.

216 The speed and duration of the motion were highly similar to that used in a previous  
217 human fMRI study (Orlov and Zohary, 2018). These authors did not employ a noise  
218 background, but we included it to have additional occlusion cues. In our conditions, the  
219 shape is perceived as moving behind the noise background, becoming partially visible  
220 through the static slit aperture. Another difference between our display and that of the  
221 human fMRI study (Orlov and Zohary, 2018) is that we had separate trials for the two motion  
222 directions, instead of a presentation of the two motion directions immediately after each  
223 other in a single trial. Unlike in the human fMRI study (Orlov and Zohary, 2018), in which the  
224 whole shapes were presented only after the slit-viewing conditions, we presented the whole  
225 shapes in the search task before the slit-viewing tasks and interleaved them with the slit-  
226 viewing conditions. This could only have increased the percept of the shape during slit-  
227 viewing. The presentation of the empty slit, well before the shape became visible, aimed to  
228 reduce potential neural responses to the onset of the slit itself by the time of shape onset.  
229 The Orlov and Zohary (2018) study employed novel, unfamiliar shapes, while our shapes  
230 were familiar to the monkey since these were presented repeatedly while searching for  
231 neurons. However, which shape was presented during slit-viewing on a particular trial was  
232 unpredictable.

233 In the slit-viewing with randomly-ordered shape fragments conditions (“random”  
234 condition), we presented frames from the slit-viewing with original shape conditions in a  
235 pseudo-random order (Figure 1C). The order of the shape parts was random, except that  
236 fragments that followed each other in the original slit-viewing condition were not allowed to  
237 be sequential in the random condition. However, we made two changes to the random-  
238 order conditions employed in a previous human fMRI study (Orlov and Zohary, 2018). First,  
239 we presented only frames with non-overlapping shape segments and these were presented  
240 for 6 successive frames (80ms) each. The latter reduced the contribution of forward and  
241 backward masking (Kovacs et al., 1995a) to the responses to the partial shape views in these  
242 random conditions, which are expected to have a strong impact when using a random  
243 ordering of the frames at the original frame rate, as in the human fMRI study (Orlov and  
244 Zohary, 2018). Thus, we presented 9 fragments in random order for 6 frames each, and  
245 these were preceded and followed by two frames of the start and end fragments,  
246 respectively, of the corresponding slit-viewing with original shape condition (Figure 1C). The  
247 total duration equaled that of the original slit-viewing condition. Second, the order of the  
248 frames was fixed across trials, allowing us to average across trials the responses using short  
249 time bins. The shape fragments of the random condition are a subset of those shown in the  
250 single frames of the original slit-viewing condition since the partial shape sections of  
251 successive frames in the latter condition partially overlapped because of the smooth motion  
252 of the shape. To control for the partial overlap of the shape fragments in the original slit-  
253 viewing and the random condition, we also made recordings in a subset of neurons with a  
254 third kind of display in which the same fragments that were presented in the random  
255 condition were shown in their correct order. This yielded a similar percept as in the  
256 corresponding original slit-viewing conditions, except for some minor jumps across the 6

257 frames long presentations. We denote these displays as “jumping” displays. Thus, in the  
258 random and jumping conditions, the same shape fragments were presented through the  
259 aperture but in a different order.

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

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

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

281 to the shape that matched the sample stimulus. The match and non-match stimuli stayed on  
282 the screen for 4 s or until the monkey made a saccade to one of the stimuli. Correct  
283 responses were rewarded with a fluid reward.

284 The monkey was trained extensively in this task with a large variety of 60 shapes.  
285 After this training, we presented old and novel shapes (30) as sample stimuli. The match and  
286 non-match stimuli could be either novel, old and novel, or both old. Using a subset of stimuli  
287 (10 old, 10 novel shapes), sample stimuli could be slit-view presentations of either moving  
288 shapes or randomly ordered snapshots of the slit-views as in the random PF condition. The  
289 different tested combinations of sample, match and non-match stimuli will be described in  
290 the Results.

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

298 *Data analysis.*

299 Responsiveness and selectivity. Because we searched for neurons using static, whole  
300 shapes with the search test, all neurons responded well to at least the best static shape  
301 condition of the subsequent slit-viewing test. We assessed significant responses of each  
302 neuron to the original slit-viewing conditions of the slit-viewing test using a three-way Split-  
303 Plot ANOVA with the repeated measure factor “epoch” (9 levels corresponding to 9 windows  
304 of 100 ms each, starting 100 ms before motion onset) and between-trial factors “shape” (3

305 levels: best, medium and worst) and “motion direction” (4 levels: leftwards (RL), rightwards  
306 (LR), upwards (DU) and downwards (UD)). We employed windows of 100 ms for the factor  
307 “epoch” since we noted during the recordings that the neurons responded during a limited  
308 period of the movie and we wished to capture such modulation of the response during the  
309 slit-viewing movie. Neurons that showed a significant effect ( $p < 0.05$ ) of the factor “epoch”  
310 or an interaction of the factors “shape” and “epoch” were considered to show a significant  
311 response to the slit-viewing stimuli.

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

324 To quantify the extent to which shape and motion direction were encoded in a  
325 separable manner, we computed for each responsive neuron a “Separability Index”. This  
326 index compares the responses to the 3 shape x 4 motion direction combinations of the  
327 “original” condition to responses predicted under the assumption that the response to each  
328 combination results from independent tuning along the shape and motion direction

329 dimensions. We followed a previously published procedure (Mysore et al., 2010) to compute  
330 this index. The mean firing rates, computed in the 800 ms long window and averaged over  
331 trials, for the 12 stimuli were tabulated in a  $3 \times 4$  response matrix ( $M$ ) with  $m$  and  $n$   
332 corresponding to the 3 shapes and the 4 motion directions, respectively. We then computed  
333 the singular value decomposition ( $M = USV'$ ) of the response matrix. The predicted response  
334 was the product of the first columns of  $U$  and  $V$  of the singular value decomposition. The  
335 Separability Index equals the squared Pearson correlation ( $r^2$ ) between the actual and  
336 predicted responses.

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

343       Direction Index =  $(\text{Response best direction} - \text{Response direction } i) / (\text{Response best}$   
344  $\text{direction} + \text{Response direction } i)$ ,  
345 with motion direction  $i$  being either the direction opposite to the best direction along the  
346 same axis or the two directions of the orthogonal axis. Thus, we computed 3 Direction  
347 Indices for each responsive neuron: one for opposite directions along the same axis, and two  
348 for directions along the axis orthogonal to the “best” axis. The index can range in principle  
349 from -1 to 1, but negative values indicate conceptually an absence of direction selectivity (as  
350 a value of zero), since responses higher than those to the best direction, measured in other  
351 trials, will reflect noise (trial-to-trial variability).

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

361        To uncover the slit-views to which the neuron responded in the “original” slit-viewing  
362        condition, we applied for each motion direction the following procedure, akin to reverse  
363        correlation. We binned the mean responses in bins of 75 ms and then assigned each binned  
364        response to the shape part that was presented 70 ms before the start of that bin. Doing so,  
365        we obtained a vector of the responses to the shape fragments during slit-viewing and that  
366        for each of the 4 directions. These vectors were then visualized on an image depicting the  
367        spatially concatenated slit-views. Further quantification was accomplished by binning the  
368        elements of the vector in 11 bins, corresponding to the shape fragments that were  
369        presented in the snapshot test.

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

375        Decoding of shape identity. We decoded the shape identity from the responses of the  
376    neurons in the “original” slit-viewing and static presentation conditions. We performed the  
377    decoding on the data pooled across monkeys. For decoding, we employed The Neural  
378    Decoding Toolbox (Meyers, 2013) and linear Support Vector Machines (SVM) as classifiers.  
379    For each neuron, 10 trials of the 15 conditions (original slit-viewing conditions (12) and the  
380    responses for static stimuli (best, medium, and worst)) were used in the analysis. We made  
381    pseudo-population responses by concatenating single-trial responses of the successively  
382    recorded neurons in a vector. Thus, each vector represented the response of the population  
383    of neurons on a trial. The responses of each neuron were z-normalized across stimulus  
384    conditions so that each neuron contributed equally. The classifier was trained using 5-fold  
385    cross-validation to control for overfitting. The reported classification accuracies are all based  
386    on zero-one loss function results. Standard deviations of classification scores were calculated  
387    across 50 cross-validated resamplings of the pseudo-population vectors. To test whether the  
388    decoding results were above chance, a permutation test (1000 permutations) with shuffled  
389    condition labels was used.

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

398 analysis allowed an assessment of the temporal specificity of shape encoding during slit-  
399 viewing.

400 Behavioral performance in DMS task. We included only unaborted trials in which the  
401 monkey made a saccade to one of the two test stimuli. Percent correct responses were  
402 computed for all trials of a test condition, as will be specified in the Results. Confidence  
403 intervals (95%) of percent correct were computed using the Binomial distribution with  
404 <https://www.graphpad.com/quickcalcs/confInterval1/>.

405 Eye movements. Eye positions along the horizontal and vertical dimensions were  
406 analyzed separately for each of the motion directions during slit-viewing. Before averaging,  
407 we subtracted for each trial the mean eye position in a 20 to 0 ms period before motion  
408 onset from the eye positions measured after motion onset. For each monkey, we averaged  
409 the baseline-subtracted eye positions per shape and motion direction for the original slit-  
410 viewing conditions. Confidence intervals (95%) were computed for each time point with  
411 bootstrap resampling.

412 Experimental Design and Statistical Analysis. We employed both parametric (ANOVA)  
413 and non-parametric tests. The factors and design of the ANOVA are described above and in  
414 the corresponding Results sections. Parametric tests were used only when no non-  
415 parametric, distribution-free tests were available. ANOVAs were performed using the R  
416 statistical software package and non-parametric tests, except noted otherwise, were  
417 performed using Matlab functions.

418

419

420 **Results**

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

433

434 Responses and selectivity in PF task.

435 We examined the responses of single body patch neurons (ASB; Figure 1B) to slit-  
436 viewing using three equally-sized silhouette shapes of animals. For every single neuron, we  
437 selected the three shapes using the responses to 70 shapes that were presented in a search  
438 test (see Materials and Methods). One of them, labeled “best”, produced the largest  
439 response of the 70 shapes, a second one, the “worst” shape, no or the weakest response,  
440 and the third shape, the “medium” one, a response in-between the best and worst shape.  
441 These 3 shapes were presented when moving behind the narrow slit in either one of two  
442 directions for each of 2 slit orientations (Figure 1) during passive fixation. In the same test,  
443 we also presented the same three shapes without motion in a large aperture of the  
444 occluding surface on a gray background.

445 We recorded the responses of 196 IT neurons, responsive to static whole shapes, in  
446 the slit-viewing test. We assessed for each neuron whether it responded significantly in at  
447 least one of the slit-viewing conditions with a three-way Split-Plot ANOVA with a repeated  
448 measure factor “epoch” (9 levels corresponding to 9 windows of 100 ms each, starting 100  
449 ms before motion onset; See Materials and Methods) and between-trial factors “shape”  
450 (best, medium and worst) and “motion direction”. The very large majority of the neurons  
451 showed a significant effect of the factor “epoch” in each of the three monkeys with an  
452 excitatory response during slit-viewing (MG: 87% (N= 108); MT: 100% (N= 63); MB: 100%  
453 (N=25)). Also, the responses of most of these neurons was modulated by shape (MG: 63%  
454 (N= 108); MT: 89% (N= 63); MB: 72% (N=25)) or motion direction ((MG: 69% (N= 98); MT:  
455 89% (N= 63); MB: 80% (N=25)).

456 Figure 2 shows the responses of a responsive single neuron to the slit-viewing  
457 conditions and the static shape presentations. As expected from the preceding Search test,  
458 the neuron produced the largest response to the selected best static shape, no response to  
459 the selected worst static shape, and an intermediate response to the medium static shape.  
460 In the slit-viewing conditions, this neuron did not respond to the slit-onset itself, which  
461 occurred 480 ms before the shape started to move. The neuron responded when the best  
462 shape was moving along the horizontal axis behind the slit, while it showed less response  
463 when the same shape was moving along the vertical axis. The neuron showed less if any  
464 response to the medium and worst shapes when these were presented during slit-viewing.  
465 Thus, for the horizontal axis slit-viewing conditions, the shape preference fitted that of the  
466 whole shape presentation. Note that the neuron responded only during a brief period of slit-  
467 viewing for a particular direction, which was a common finding in our sample of neurons.  
468 The timing of this responsive period differed between the two horizontal directions. Noted

469 that the parts that were presented at the beginning of one motion direction occurred at the  
470 end of the other direction for the same slit orientation. There was no evidence of temporal  
471 integration of the responses during the slit-viewing, which was typical for our sample of  
472 neurons.

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

481 Other neurons responded with similar peak firing rates in original and random slit-  
482 viewing conditions. One example (Figure 3A) of such neurons responded for the vertical  
483 motion slit-viewing conditions of the best and medium shape. It also responded for those  
484 directions in the random conditions of the medium (but not best) shape. The neuron of  
485 Figure 3B responded in the horizontal slit-viewing conditions of the best shape, and this for  
486 both original and random slit-viewing conditions. Note that, as for the neuron of Figure 2,  
487 both neurons responded during a brief period of the slit-viewing movie, and this for both the  
488 original and random conditions. Figure 3C shows a neuron that responded somewhat longer  
489 during the slit-viewing, but the period during which it responded during slit-viewing  
490 depended on motion direction and it showed strong selectivity for the motion axis. None of  
491 the neurons in Figure 3 showed evidence of temporal integration of the responses during  
492 slit-viewing.

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

506        The response during slit-viewing was greater for the best compared to the worst  
507    shape, with an intermediate response for the medium shape. Note that the best, medium,  
508    and worst shapes were defined based on the response to static presentations of the whole  
509    shape. Thus, the shape preference of the population response was invariant to the viewing  
510    conditions, although the responses to the static presentations of the whole shapes were  
511    markedly greater than the average responses during slit-viewing of the same shapes (Figure  
512    4A). The responses for the random conditions tended to be smaller than those for the  
513    original slit-viewing conditions, but even for the random conditions, the population  
514    responses were larger for the best compared to the worst shape conditions.

515        To assess the statistical significance of the effect of shape and the difference  
516    between random and original slit-viewing conditions, we computed for each neuron the

517 response for each slit-viewing condition (3 shapes x 4 directions x random versus original  
518 conditions) using an analysis window of 800 ms that started 50 ms after motion onset (the  
519 duration of slit-viewing of the shape was 773 ms). We performed a three-way repeated  
520 measure ANOVA of the responses of the 185 neurons with repeated factors shape, motion  
521 direction, and original versus random slit-viewing conditions. The factor shape was highly  
522 significant ( $F(1.76, 664.5) = 116.97; p = 1.8 \times 10^{-35}$ ; Greenhouse-Geisser (sphericity)  
523 corrected): the mean response for both the original and random conditions was the largest  
524 for the best shape (defined using static whole-shape presentations), intermediate for the  
525 medium shape and the smallest for the worst shape (Figure 4B). This difference in mean  
526 responses across the shape was significant in each monkey. There was a significant effect of  
527 motion direction ( $F(1.5, 828) = 5.13; p = 1.2 \times 10^{-2}$ ; Greenhouse-Geisser corrected) with on  
528 average stronger responses for the horizontal axis (vertical slit) than vertical axis directions  
529 (horizontal slit; Figure 4B). However, this effect was absent in monkey MT. Furthermore, the  
530 factor motion axis is confounded with a difference in visual field location of the slits and  
531 hence this effect is difficult to interpret. Mean responses were significantly greater for the  
532 original compared to the random slit-viewing conditions ( $F(1, 184) = 51.69; p = 1.6 \times 10^{-11}$ ;  
533 Figure 4B) and this effect was significant in each monkey.

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

541 fragments were the same as in the random condition but were shown in the same order as  
542 in the corresponding original condition (“jumping conditions”). As in the random conditions,  
543 each shape part was presented for 80 ms in these jumping conditions. The mean responses  
544 in the jumping condition were in-between those of the original and random condition  
545 (Figure 4C) and this trend was observed in each monkey. Performing a three-way repeated  
546 measure ANOVA of the responses in the random and jumping conditions of the 126 neurons  
547 with repeated factors shape, motion direction, and jumping versus random conditions  
548 showed a significant effect of the latter factor ( $F(1, 125) = 9.08; p= 3.12* 10^{-3}$ ). At a  
549 superficial level, the greater response to the jumping compared to the random order  
550 condition might be taken as evidence that there is temporal integration of a shape during  
551 slit-viewing. However, based on the data we will present below, we prefer an alternative  
552 interpretation following the observation that spatially neighboring fragments of a natural  
553 shape can contain similar features to which a neuron responds. The latter will result in a  
554 longer and stronger response to those successive views, similar to the effect on the response  
555 of increasing stimulus duration. When, as in the random condition, these shape features are  
556 shown temporally further apart, interleaved with other shape parts, responses are expected  
557 to be smaller.

558

559 Decoding of shape identity from responses in slit-viewing conditions.

560 A human fMRI study (Orlov and Zohary, 2018) showed that shape identity could be  
561 decoded from the multivariate BOLD response in LOC during slit-viewing. Importantly, they  
562 reported generalization of classification across slit orientations and for slit-viewing and the  
563 static whole shape presentations. The hemodynamic response is sluggish, integrating neural  
564 activity across time. Although single IT neurons showed no evidence of temporal integration

565 of the whole shape during slit-viewing, the shape preference of the average population  
566 response during slit-viewing, temporally integrated by averaging in the 800 ms analysis  
567 window, was the same as for the static shape presentations (Figure 4). This suggests that it is  
568 possible to decode shape identity from the slit-viewing response and that there is a  
569 generalization of shape identity classification across slit-viewing conditions and for the static  
570 and slit-view presentations. We examined this by training a linear classifier (SVM; see  
571 Materials and Methods) to classify shape using as input single-trial pseudo-population  
572 response vectors that consisted of the single-trial responses, averaged in the 800 ms  
573 window, of the recorded neurons. We trained the classifier using the data of one of the 5  
574 conditions (static, LR, RL, UD, and DU slit-viewing) and then tested classification for the  
575 same, trained, condition (5-fold cross-validation) or the untrained 4 other conditions (cross-  
576 condition classification).

577 When testing and training were performed using data of the same condition,  
578 classification of shape identity was close to or at the ceiling level (hatched bars in Figures 5A-  
579 E) for both static (S) and slit-viewing conditions. This demonstrates that shape identity can  
580 be decoded reliably from the temporally integrated responses of the recorded sample of  
581 neurons during slit-viewing. For cross-condition classification (blue bars in Figures 5A-E), the  
582 amount of generalization of classification depended on the trained and tested conditions.  
583 We observed excellent generalization across orthogonal motion directions of the slit-viewing  
584 conditions (e.g. train LR, test RL). The classification scores dropped to about 60% correct, but  
585 were still significantly above chance (33.3%), when trained and tested motion axes differed  
586 (e.g. train LR, test UD), except when training LR and testing DU (Figure 5B). This  
587 demonstrates the generalization of shape classification across slit orientation when  
588 integrating the responses across time. Training the classifier with the responses recorded

589 during slit-viewing, we obtained also well above-chance classification of the static shape  
590 (Figures 5B-E). In fact, the cross-condition test classification scores for the whole shape were  
591 similar to those obtained when trained and tested slit orientations were orthogonal.  
592 However, training the classifier with the responses to the static whole shape yielded chance  
593 classification scores when testing slit-view responses (Figure 5A). Such asymmetry of cross-  
594 condition classification has been observed before when conditions differ markedly in  
595 response strength and signal-to-noise (SNR) ratio (Van den Hurk and Op de Beeck, 2019), as  
596 is also the case here (Figure 4A). Modeling, using linear SVM, has demonstrated that having  
597 a low SNR condition A and a high SNR condition B results in better generalization when  
598 testing B after training A than vice versa (Van den Hurk and Op de Beeck, 2019). Also, when  
599 only a subset of the informative neurons in condition 2 contains an informative signal for the  
600 classifier in condition 1, generalization will be better when training condition 1 and testing 2  
601 than vice versa (Van den Hurk and Op de Beeck, 2019). Both factors can explain the  
602 asymmetry in generalization performance seen in our data since shape selectivity of the  
603 single neurons was more robust for the static whole shape presentation compared to the  
604 slit-viewing conditions.

605 In sum, the cross-condition classification data show evidence for generalization  
606 across motion direction, slit orientation, and whole-shape versus slit-view presentations  
607 when integrating the response throughout the slit-viewing period. These results of monkey  
608 IT single-unit data are in line with the generalization of shape identity classification obtained  
609 for BOLD activation patterns in human LOC (Orlov and Zohary, 2018). Since the BOLD  
610 hemodynamic response function causes integration of the neural responses across time it  
611 can produce similar results as we obtained here by temporal integration of the spiking  
612 activity of single neurons. However, the underlying dynamics of the stimulus representation

613 during slit-viewing are lost when temporally integrating neural responses. To capture the  
614 dynamics of the shape representation during slit-viewing, we decoded shape identity with  
615 classifiers that were trained and tested with brief 100 ms analysis windows. We performed  
616 classification of shape identity for different training and testing periods and this when  
617 trained and tested conditions were identical or differed. Figure 5F shows the classification  
618 scores for all possible training-testing time combinations (cross-temporal decoding), starting  
619 200 ms before motion onset and ending 227 ms after motion offset, and this for all the 25  
620 possible combinations of the trained and tested stimulus conditions. When trained and  
621 tested conditions were identical (panels along the (top) left to the right diagonal in Figure  
622 5F), the classification accuracy dropped markedly when training and testing differed by more  
623 than 200 ms for all slit-viewing conditions. This is clearly shown by the reddish left diagonal  
624 band in the cross-temporal decoding plots. This temporally-specific code contrasted with the  
625 more stationary one for the static, whole-shape presentations. Importantly, the cross-  
626 temporal decoding plots for opponent trained and tested motion directions showed also a  
627 clear diagonal band but from right to left. Thus, the generalization of shape classification  
628 across opponent motion directions is highly temporally-specific. The mirror-symmetry of the  
629 cross-temporal decoding plots for identical versus opponent trained and tested motion  
630 directions suggests that the neurons responded to shape fragments ("effective shape  
631 fragments") that were visible at a particular moment of the slit-viewing: an effective shape  
632 fragment that is visible e.g. at the beginning of the slit-viewing for one motion direction (e.g.  
633 LR) will be visible at the end of the slit-viewing for the opponent motion direction (e.g. RL).  
634 When trained and tested conditions consisted of different motion axes (horizontal versus  
635 vertical), there will not be a correlation of effective shape fragments across time between  
636 the axes, which results in more diffuse, less organized cross-temporal decoding plots for the

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

646

647 Responses to shape fragments during slit-viewing and static presentation.

648 The cross-temporal decoding analysis (Figure 5F) suggests that the responses during  
649 slit-viewing are driven by effective shape fragments that become briefly visible. This would  
650 imply that the neurons respond to the thin shape strip that is visible through the narrow slit.  
651 We tested this directly by presenting 11 snapshots of the slit-viewing movie to a sample of  
652 neurons (in two monkeys) that were also tested during slit-viewing. The snapshots were  
653 presented briefly for 80 ms, each preceded and succeeded by the background noise pattern  
654 with the empty slit (see Materials and Methods). An example neuron tested in this snapshot  
655 test is shown in Figure 6B. It responded selectivity to shape features related to the arms of  
656 an ape silhouette. To relate the responses in the snapshot test to those obtained during slit-  
657 viewing, we employed a method akin to reverse correlation to compute the responses to  
658 individual shape fragments during slit-viewing (see Materials and Methods). This procedure  
659 yields a response for each frame of the slit-viewing movie, which can be visualized as a  
660 shape response plot in which the response to an individual shape fragment is indicated by a

661 color code. This is illustrated for the same neuron in Figure 6A for both horizontal motion  
662 directions. This neuron responded for two brief periods during slit-viewing for both motion  
663 directions. Considering the typical response latency of IT neurons, the neuron responded  
664 during slit-viewing to parts of the arm of the ape silhouette. As shown in Figure 6B, the  
665 effective shape fragments obtained using reverse correlation of the slit-view data of each  
666 motion direction corresponded to those revealed by the snapshot test. Note that when  
667 comparing the snapshot responses to the reverse-correlated slit-viewing responses we  
668 reversed the plotted order of the fragments of one of the motion directions, so that shape  
669 fragments corresponded in the shape response plots of the two motion directions. Other  
670 examples of shape response plots for both motion directions and the corresponding  
671 snapshot test plots are shown in Figure 6D.

672 To quantify the correspondence between the responses during slit-viewing and in the  
673 snapshot test, we binned the responses of the shape response plots for the slit-viewing  
674 conditions in 11 bins and computed for each motion direction the Pearson correlation  
675 coefficient between the thus obtained responses for the slit-viewing and those of the  
676 snapshot test. For the neuron of Figure 6B, the correlation coefficients were close to 1,  
677 demonstrating the excellent fit between the responses to shape fragments during slit-  
678 viewing and in the snapshot test. We computed this correlation for all the snapshot test –  
679 slit-viewing combinations for which there was a significant response during slit-viewing  
680 (significance tested with ANOVA). A total of 45 and 15 neurons were tested in monkey MG  
681 and MT, respectively, and in 33 neurons we had snapshot tests for both motion axes. Figure  
682 6C shows the distribution of the correlation coefficients for each monkey separately, which  
683 was shifted towards positive values. The median correlation coefficients were significantly  
684 greater than 0 in each monkey (Wilcoxon test; MG: median=0.48; p = 9. 9\* 10<sup>-27</sup>; MT:

685 median=0.40;  $p = 1.6 \times 10^{-13}$ ), indicating that responses during slit-viewing and the static  
686 snapshot test are related. Thus, the observation that the single units responded during only  
687 brief periods of the slit-viewing can be explained by selective responses to effective shape  
688 fragments that were revealed by the slit during these moments.

689

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

691 The single unit IT data reported above showed no evidence of temporal integration of  
692 the whole shape during slit-viewing. This raises the question of whether macaques perceive  
693 a shape during slit-viewing, as humans do. To examine this, we trained monkey MG after the  
694 above-reported recordings in a DMS task in which he had to match a shape, presented  
695 during slit-viewing, and a static presentation of the same shape (Figure 1D). Trials in which  
696 the sample stimulus was a whole shape were interleaved with slit-viewing samples. After  
697 training with a pool of 60 shapes, the monkey was tested with various combinations of the  
698 match and non-match stimuli, including novel shapes that were not employed during the  
699 training period. Figure 7A summarizes the performance of the monkey for combinations of  
700 novel and old stimuli, presented either as sample or test stimuli. The old and novel stimulus  
701 sets consisted of 10 shapes each, and stimuli were randomly interleaved across trials. The  
702 number of sample presentations of a particular shape varied between 8 and 11. Although  
703 the DMS performance was greater when the sample stimulus was a static whole shape (82%  
704 correct; 95% confidence interval (CI)= 77- 88), the behavioral performance (70% correct; CI =  
705 67 - 73) when the sample shape was presented during slit-viewing was well above chance (=  
706 50% correct). There was no evidence of a difference in the performance between the novel  
707 and familiar, old shapes. Performance was highly similar for vertical (mean = 69%; CI = 62 -  
708 75) and horizontal slit orientations (mean = 70%; CI = 63 - 76). Analyzing the performance for

709 first trial presentations of a shape during slit-viewing also showed above-chance  
710 performance (67% correct; CI = 61 - 72), demonstrating that the above-chance performance  
711 for the slit-viewing sample conditions did not result from paired-associate learning of shape  
712 fragment samples and whole-shape match stimuli.

713 One could argue that the monkey employed isolated shape fragments but no  
714 integrated whole-shape percept to match the slit-views and whole shapes. To examine this  
715 possibility, we presented trials in which the shape fragments were presented either in their  
716 correct order or in random order, as in our random slit-viewing conditions in the single-unit  
717 recording experiment. The original whole shape and a whole “random” shape, which was a  
718 spatial concatenation of the shape parts following the order of the random slit-views (as in  
719 Figure 1C), served as the match and nonmatch stimuli. We reasoned that if isolated shape  
720 features were driving the performance of the monkey, the monkey should show poor  
721 performance when having to choose between these match and nonmatch stimuli since both  
722 contained the same shape fragments as the sample stimulus. When the sample stimulus was  
723 an original slit-view, the performance of the monkey was 67% correct (CI = 63 - 70), which  
724 was well above chance level. This suggests that the monkey is not merely relying on isolated  
725 shape fragments when matching slit-views and test stimuli. When the sample stimulus was a  
726 random slit-viewing stimulus, the performance of the monkey dropped to 46% (CI = 43 - 50),  
727 which was statistically not different from chance. However, when the random stimulus  
728 configuration was shown as a static shape and the match and non-match stimuli were  
729 original and random whole shape configurations, the performance was 74% (CI = 67 – 80),  
730 which was highly similar to the performance when the original, whole-shape was the sample  
731 stimulus (73%; CI = 66 - 79). Thus, although the monkey was able to match original and  
732 random whole-shape configurations, he was unable to match random-slit views. Such poor

733 performance was also present when the random slit-views served as sample stimuli and  
734 both match and nonmatch stimuli were random shape configurations (mean = 54% correct;  
735 CI = 49 - 59). We attribute the chance performance for random slit-views samples to the  
736 difficulty of temporally integrating the randomly ordered shape parts, because of the large  
737 spatio-temporal discontinuities between successive fragments, into a single shape percept.  
738 These behavioral data support the presence of anorthoscopic shape perception in our slit-  
739 viewing conditions in macaques.

740

741 Single unit responses during slit-viewing in DMS and passive fixation tasks.

742 To examine the possibility that temporal shape integration during slit-viewing at the  
743 IT single-unit level would occur when the monkey is attending the stimulus, we recorded 20  
744 responsive neurons while MG was performing the DMS task using slit-views and static whole  
745 shapes as sample stimuli. The sample stimuli in the DMS task were 3 shapes that were  
746 selected anew for each neuron based on the responses in the search test in which 70 shapes  
747 were presented during passive fixation. After the recordings during the DMS task, we also  
748 recorded an additional 20 responsive neurons in MG while he was performing the same slit-  
749 viewing test with passive fixation as before the DMS training. To compare quantitatively the  
750 three samples of neurons, those recorded during PF, while performing the DMS task, and  
751 during post-DMS PF, we computed five response property indices for each neuron using data  
752 for the original shape slit-viewing conditions. The first three indices quantified the effect of  
753 motion direction on the responses during slit-viewing of the best shape. The first of these  
754 Direction Indices compared the response to the best motion direction and its opponent  
755 direction for the same motion axis (see Materials and Methods). Note that for each neuron  
756 its best direction was determined in independent trials, explaining the presence of negative

757 Direction Indices. For the three samples of neurons of MG, the median Direction Indices for  
758 same axis directions were low (Figure 8B) and similar to those obtained in the other two  
759 monkeys during passive fixation (Figure 8A). The low median values (approximately 0.1, or a  
760 22% difference) agree with the finding that the responses during slit-viewing are driven by  
761 effective shape fragments, which are identical for the two directions of the same axis.  
762 However, for all 5 samples, the same-axis Direction Index was positive, and significantly  
763 greater than zero in each monkey (MG:  $p = 1.2 \times 10^{-5}$ ; MT:  $p = 3.7 \times 10^{-3}$ ; MB:  $p = 5.8 \times 10^{-5}$ ;  
764 Wilcoxon test), showing an albeit weak influence of motion direction on the responses. This  
765 agrees with previous studies that showed stimulation history effects on IT responses (see  
766 Discussion).

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

779 We also quantified for each neuron the period during which the neuron responded  
780 during slit-viewing by measuring the duration of its response at half-height (see Materials

781 and Methods). This peak duration estimation was performed for the slit-viewing condition  
782 that produced the maximum response of the neuron. One would predict longer peak  
783 durations after DMS training or during DMS task performance when this increased temporal  
784 shape integration. However, the opposite trend was present (Figure 8F; PF versus DMS:  $p =$   
785 0.0035; PF versus post-DMS:  $p = 0.0018$ ; Wilcoxon rank sum test), suggesting that temporal  
786 shape integration did not increase during or after the DMS task. Note that the median peak  
787 durations during the DMS task and post-DMS in MG are similar to those obtained during the  
788 PF task in the other two monkeys (Figure 8E).

789 To assess the extent to which shape and motion direction were encoded in a  
790 separable manner, we computed for each neuron the Separability Index (see Materials and  
791 Methods). A high Separability Index indicates that one can predict the responses in a slit-  
792 viewing condition by knowing the responses to the shape, irrespectively of motion direction,  
793 and to the motion direction, irrespectively of shape. High separability will ensure invariant  
794 decoding of the shape, irrespectively of motion direction (Li et al., 2009). The Separability  
795 Indices were computed after integrating the responses during slit-viewing using the 800 ms  
796 analysis window. We found that the median Separability Index was significantly greater  
797 when the monkey was performing the DMS task compared to the passive fixation before the  
798 training ( $p = 5.3 * 10^{-3}$ ; Wilcoxon rank sum test; Figure 8G). This was not due to the difference  
799 between the two tasks, since the neurons recorded post-training during passive fixation  
800 showed also a greater median Separability Index than the sample of neurons recorded  
801 before the DMS training ( $p = 7.9 * 10^{-4}$ ; Wilcoxon rank sum test; Figure 8G). One possible  
802 explanation for the increased Separability Index after DMS training is that the shape  
803 selectivity of the neurons was greater post-training. To address this possibility, we computed  
804 for each neuron a Shape Selectivity index, being the response to the best static whole-shape

805 minus the response to the worst static whole-shape, divided by the sum of the responses to  
806 both shapes. The median Shape Selectivity index was indeed significantly greater after  
807 (median index during DMS task: 0.71; post-DMS: 0.9) compared to before DMS training  
808 (median: 0.57;  $p = 1.2 * 10^{-6}$ ; Wilcoxon rank sum test; data of DMS and post-DMS tasks  
809 pooled). Thus, the shape selectivity was higher for the sample of neurons recorded after the  
810 DMS training, which can explain their higher Separability Indices.

811

812       **Eye movements.**

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

818

819       **Discussion**

820       We recorded the responses of body patch (ASB) neurons to moving shapes that were  
821 only partially visible through a static narrow slit. Although only a small fragment of the shape  
822 was revealed through the slit at a single moment in time, the population of IT neurons  
823 signaled shape identity by their response when that was cumulated across the viewing  
824 period. The shape preference for the slit-viewing conditions was the same as for static  
825 whole-shape presentations. However, by analyzing the responses on a finer time scale and  
826 comparing responses between motion directions, we showed that IT neurons responded to  
827 particular shape features that were visible through the slit. Furthermore, the responses  
828 during slit-viewing were predicted by the responses of the same neuron to static

829 presentations of shape fragments as revealed through the slit. In this IT body patch, we  
830 found no evidence for temporal integration of the sequentially revealed shape parts into a  
831 coherent shape representation, neither at the single unit nor the population response level.  
832 Qualitatively identical response patterns were present when a monkey was matching slit-  
833 views of a shape to static whole-shape presentations and thus attending the slit-views.  
834 These data suggest that although the temporally integrated response of macaque IT neurons  
835 can signal shape identity under slit-viewing, the temporal integration needed for the  
836 formation of a whole-shape percept occurs in other areas, perhaps downstream to IT.

837 The following observations lead to the conclusion that single IT neurons responded to  
838 partial shape views, but did not integrate the whole shape, during slit-viewing. First, single IT  
839 neurons responded only during particular periods of the slit-viewing movie and these  
840 corresponded to periods in which effective shape fragments were displayed as determined  
841 in an independent test with static snapshots of the movie. Second, shape decoding during  
842 slit-viewing was restricted in time and generalized across opposite motion directions for the  
843 different periods in which the same shape fragment was presented for the two directions.  
844 Third, the generalization of shape decoding was less across axes than across directions of the  
845 same axis, which is expected when the neurons respond to shape fragments because shape  
846 features overlap less for different slit orientations. However, whole-shape representations  
847 are expected to be invariant to slit orientation. Furthermore, shape selectivity did not  
848 consistently increase during slit-viewing. Fourth, the average shape preference was the same  
849 for the random and original slit-viewing conditions, supporting the conclusion that the  
850 neurons represent shape fragments.

851 The response to the slit-views could differ, albeit weakly, between opposite motion  
852 directions, although shape fragments were equal for the two directions. This is not surprising

853 since the response to a particular shape fragment will also be determined by the response  
854 selectivity of the neuron for preceding shape features, which will differ between directions.  
855 Indeed, responses of IT neurons depend on stimulation history and preceding stimuli, e.g.  
856 adaptation (Vogels, 2016) and shape sequence effects as observed for forward versus  
857 backward walking sequences (Vangeneugden et al., 2011). Although these effects of  
858 preceding stimulation history and the motion direction sensitivity found in the present study  
859 show that IT neurons show temporal integration of preceding and current stimulation, this  
860 should be distinguished from temporal integration of dynamic partial shape views for the  
861 formation of a whole-shape percept during slit-viewing. The latter requires temporal  
862 integration and maintenance of shape information across slit-views, taking into account the  
863 relative location of the shape fragments based on the velocity of the features in the slit  
864 during stimulation (Öğmen and Herzog, 2016). Our results are in line with previous estimates  
865 of a temporal integration duration of 100-120 ms in the macaque rostral STS for visual action  
866 sequences (Singer and Sheinberg, 2010; Vangeneugden et al., 2011), which is shorter than  
867 required for integration of slit-views into a whole-shape percept.

868       Although the recorded neurons did not temporally integrate the slit-views, a linear  
869 classifier could classify the shapes when it had as input the responses averaged over the  
870 whole slit-viewing period. Furthermore, for this averaged response, a generalization of  
871 shape classification occurred across motion directions, across motion axes, and for static  
872 whole-shape and slit-view presentations. These results agree with a human fMRI study in  
873 which generalization of shape classification across slit-viewing axes and for static whole-  
874 shape was shown with MVPA in ventral stream visual areas (Orlov and Zohary, 2018). Note  
875 that because of the temporally coarse hemodynamic response, the BOLD response amounts  
876 to using a long window in which responses will be temporally integrated, similar to what we

877 did when averaging the responses over the slit-viewing period. However, our results show  
878 that such generalization of classification using the integrated response cannot be used as  
879 evidence for a temporally integrated whole-shape representation at the level of single  
880 neurons. In contrast to the presence of shape selectivity in the random slit-viewing  
881 conditions in our macaque data, classification of shape based on random slit-view  
882 presentations was at chance level in the human fMRI study (Orlov and Zohary, 2018).  
883 However they randomized the frames at a 60 Hz rate in their random condition. This  
884 amounts to a rapid serial presentation of randomly ordered shape fragments at 60 Hz. In  
885 macaque STS, such rapid serial presentation decreases stimulus selective responses (Keysers  
886 et al., 2005; De et al., 2007) because of forward and backward masking, and we expect the  
887 same reduction in human ventral stream areas. This may explain why shape classification  
888 with MVPA of LOC activations was at chance level in the random slit-viewing conditions  
889 (Orlov and Zohary, 2018). As noted above, moving shapes can produce stronger responses  
890 because of the similarity of the nearby stimulus features during slit-viewing.

891 One behavioral study suggested poor spatio-temporal integration during slit-viewing  
892 in apes compared to humans (Imura and Tomonaga, 2013), but such quantitative  
893 interspecies comparisons are difficult to interpret since non-perceptual, cognitive  
894 differences between species can affect visual task performance. What that study did show is  
895 that chimpanzees perform better than the chance level when matching dynamic slit-views to  
896 static whole-shape outlines. We found a similar result here in a macaque, suggesting that  
897 nonhuman primates show AP. This also suggests that the lack of a temporally integrated  
898 whole-shape representation at the level of macaque IT neurons during slit-viewing is not  
899 because monkeys do not show AP. The single-unit properties were similar when the monkey  
900 was performing matching of slit-views and thus attending the slit-view presentations. This

901 observation together with the presence of fMRI activations in LOC during the performance  
902 of an orthogonal task during slit-viewing (Yin et al., 2002), suggests that the use of a passive  
903 fixation task in most of our recordings cannot explain the lack of temporal whole-shape  
904 integration for the IT responses.

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

917 Although our findings suggest that single IT neurons do not form whole-shape  
918 representations during slit-viewing, their responses, integrated across slit-viewing, contain  
919 sufficient information for shape identification. This begs the question of where such  
920 temporal integration occurs and even whether IT plays a role in building a whole-shape  
921 representation during slit-viewing. Psychophysical studies show that AP depends on the  
922 estimation of shape velocity (Morgan et al., 1982; Shimojo and Richards, 1986) which may  
923 imply the contribution of dorsal visual areas in the formation of a shape percept during slit-  
924 viewing. Indeed, human fMRI studies showed dorsal visual area (e.g. hMT/V5) activations

925 during slit-viewing, although these were less than in ventral areas (Yin et al., 2002; Orlov and  
926 Zohary, 2018). An area that underlies spatio-temporal shape integration during slit-viewing  
927 needs to be able to maintain and update shape information during viewing in  
928 “nonretinotopic” memory (Öğmen and Herzog, 2016), which requires temporal integration  
929 with a long time constant. Because temporal integration constants increase along the  
930 cortical hierarchy (Hasson et al., 2008; Murray et al., 2014; Spitmaan et al., 2020), it is  
931 possible that higher-order regions, such as the prefrontal cortex or more anterior in  
932 temporal cortex than ASB, might underlie whole-shape formation during slit-viewing, but  
933 testing this requires further work. Finally, we note that current computational models of the  
934 ventral visual stream (Kar et al., 2019) can accommodate the observed responses in  
935 macaque IT during slit-viewing. However, to explain AP computational models of visual  
936 recognition need to be augmented.

937

- 938           **References:**
- 939
- 940 Bao P, She L, McGill M, Tsao DY (2020) A map of object space in primate inferotemporal cortex.  
941           Nature 583:103-108.
- 942 De Baene W, Premereur E, Vogels R (2007) Properties of shape tuning of macaque inferior temporal  
943           neurons examined using rapid serial visual presentation. JNeurophysiol 97:2900-2916.
- 944 Denys K, Vanduffel W, Fize D, Nelissen K, Peuskens H, Van Essen D, Orban GA (2004) The processing  
945           of visual shape in the cerebral cortex of human and nonhuman primates: A functional  
946           magnetic resonance imaging study. Journal of Neuroscience 24:2551-2565.
- 947 Ekstrom LB, Roelfsema PR, Arsenault JT, Bonmassar G, Vanduffel W (2008) Bottom-up dependent  
948           gating of frontal signals in early visual cortex. Science 321:414-417.
- 949 Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N (2008) A hierarchy of temporal receptive windows in  
950           human cortex. J Neurosci 28:2539-2550.
- 951 Imura T, Tomonaga M (2013) Differences between chimpanzees and humans in visual temporal  
952           integration. Sci Rep 3:3256.
- 953 Issa EB, Dicarlo JJ (2012) Precedence of the Eye Region in Neural Processing of Faces. Journal of  
954           Neuroscience 32:16666-16682.
- 955 Kar K, Kubilius J, Schmidt K, Issa EB, DiCarlo JJ (2019) Evidence that recurrent circuits are critical to  
956           the ventral stream's execution of core object recognition behavior. Nat Neurosci 22:974-983.
- 957 Keysers C, Xiao DK, Foldiak P, Perrett DI (2005) Out of sight but not out of mind: the neurophysiology  
958           of iconic memory in the superior temporal sulcus. Cogn Neuropsychol 22:316-332.
- 959 Kovacs G, Vogels R, Orban GA (1995a) Cortical correlate of pattern backward masking.  
960           ProcNatlAcadSciUSA 92:5587-5591.
- 961 Kovacs G, Vogels R, Orban GA (1995b) Selectivity of macaque inferior temporal neurons for partially  
962           occluded shapes. JNeurosci 15:1984-1997.
- 963 Kriegeskorte N (2015) Deep Neural Networks: A New Framework for Modeling Biological Vision and  
964           Brain Information Processing. Annu Rev Vis Sci 1:417-446.
- 965 Kumar S, Popivanov ID, Vogels R (2017) Transformation of Visual Representations Across Ventral  
966           Stream Body-selective Patches. Cereb Cortex:1-15.
- 967 Li N, Cox DD, Zoccolan D, Dicarlo JJ (2009) What response properties do individual neurons need to  
968           underlie position and clutter "invariant" object recognition? JNeurophysiol 102:360-376.
- 969 McCloskey M, Watkins MJ (1978) The seeing-more-than-is-there phenomenon: implications for the  
970           locus of iconic storage. J Exp Psychol Hum Percept Perform 4:553-564.
- 971 Meyers EM (2013) The neural decoding toolbox. Front Neuroinform 7:8.
- 972 Morgan MJ, Findlay JM, Watt RJ (1982) Aperture viewing: a review and a synthesis. Q J Exp Psychol A  
973           34:211-233.
- 974 Murray JD, Bernacchia A, Freedman DJ, Romo R, Wallis JD, Cai X, Padoa-Schioppa C, Pasternak T, Seo  
975           H, Lee D, Wang XJ (2014) A hierarchy of intrinsic timescales across primate cortex. Nat  
976           Neurosci 17:1661-1663.
- 977 Mysore SG, Vogels R, Raiguel SE, Todd JT, Orban GA (2010) The selectivity of neurons in the macaque  
978           fundus of the superior temporal area for three-dimensional structure from motion. JNeurosci  
979           30:15491-15508.
- 980 Orlov T, Zohary E (2018) Object Representations in Human Visual Cortex Formed Through Temporal  
981           Integration of Dynamic Partial Shape Views. J Neurosci 38:659-678.
- 982 Parks TE (1965) Post-retinal visual storage. Am J Psychol 78:145-147.
- 983 Popivanov ID, Schyns PG, Vogels R (2016) Stimulus features coded by single neurons of a macaque  
984           body category selective patch. ProcNatlAcadSciUSA 113:E2450-E2459.
- 985 Popivanov ID, Jastorff J, Vanduffel W, Vogels R (2012) Stimulus representations in body-selective  
986           regions of the macaque cortex assessed with event-related fMRI. Neuroimage 63:723-741.
- 987 Popivanov ID, Jastorff J, Vanduffel W, Vogels R (2014) Heterogeneous Single-Unit Selectivity in an  
988           fMRI-Defined Body-Selective Patch. Journal of Neuroscience 34:95-111.

- 989 Reichert C, Fendrich R, Bernarding J, Tempelmann C, Hinrichs H, Rieger JW (2014) Online tracking of  
990 the contents of conscious perception using real-time fMRI. *Front Neurosci* 8:116.  
991 Rock I, Sigman E (1973) Intelligence factors in the perception of form through a moving slit.  
992 *Perception* 2:357-369.  
993 Sawamura H, Orban GA, Vogels R (2006) Selectivity of neuronal adaptation does not match response  
994 selectivity: a single-cell study of the fMRI adaptation paradigm. *Neuron* 49:307-318.  
995 Shimojo S, Richards W (1986) "Seeing" shapes that are almost totally occluded: a new look at Parks's  
996 camel. *Percept Psychophys* 39:418-426.  
997 Singer JM, Sheinberg DL (2010) Temporal cortex neurons encode articulated actions as slow  
998 sequences of integrated poses. *JNeurosci* 30:3133-3145.  
999 Spitmaan M, Seo H, Lee D, Soltani A (2020) Multiple timescales of neural dynamics and integration of  
1000 task-relevant signals across cortex. *Proc Natl Acad Sci U S A* 117:22522-22531.  
1001 Taubert J, Van Belle G, Vanduffel W, Rossion B, Vogels R (2015) Neural Correlate of the Thatcher Face  
1002 Illusion in a Monkey Face-Selective Patch. *J Neurosci* 35:9872-9878.  
1003 Tsao DY, Freiwald WA, Knutson TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque  
1004 cerebral cortex. *Nat Neurosci* 6:989-995.  
1005 Van den Hurk J, Op de Beeck HP (2019) Generalization asymmetry in multivariate cross-classification:  
1006 When representation A generalizes better to representation B than B to A. In. bioRxiv.  
1007 Vanduffel W, Fize D, Mandeville JB, Nelissen K, Van Hecke P, Rosen BR, Tootell RBH, Orban GA (2001)  
1008 Visual motion processing investigated using contrast agent-enhanced fMRI in awake  
1009 behaving monkeys. *Neuron* 32:565-577.  
1010 Vangeneugden J, De Maziere PA, Van Hulle MM, Jaeggli T, Van GL, Vogels R (2011) Distinct  
1011 mechanisms for coding of visual actions in macaque temporal cortex. *JNeurosci* 31:385-401.  
1012 Vinken K, Op de Beeck HP, Vogels R (2018) Face Repetition Probability Does Not Affect Repetition  
1013 Suppression in Macaque Inferotemporal Cortex. *J Neurosci* 38:7492-7504.  
1014 Vogels R (2016) Sources of adaptation of inferior temporal cortical responses. *Cortex* 80:185-195.  
1015 Yamins DL, DiCarlo JJ (2016) Using goal-driven deep learning models to understand sensory cortex.  
1016 *Nat Neurosci* 19:356-365.  
1017 Yin C, Shimojo S, Moore C, Engel SA (2002) Dynamic shape integration in extrastriate cortex. *Curr Biol*  
1018 12:1379-1385.  
1019 Öğmen H, Herzog MH (2016) A New Conceptualization of Human Visual Sensory-Memory. *Front*  
1020 *Psychol* 7:830.  
1021  
1022

1023           **Figure Legends**

1024           **Figure 1.** Stimuli and tasks. **A.** Examples of shapes used in the electrophysiological  
1025 experiments. **B.** Coronal MRI sections showing the location of the body patch ASB (indicated  
1026 by the oval outline) in monkey MG. The fMRI activations of the body patch localizer are  
1027 shown in hot colors (threshold  $t > 5$ ). **C.** Slit-viewing test during passive fixation (PF).  
1028 Following a 300 ms fixation period, either an empty horizontal gray slit (panel 1), a vertical  
1029 gray slit (panel 2), or a gray square aperture that included a static shape was presented in  
1030 the noise background. The static shape configuration was presented for 1333 ms, equal to  
1031 the duration of the slit. After 480 ms of empty slit presentation, shape passed behind the  
1032 slit. The movie displaying the shape fragments lasted 773 ms. After the presentation of the  
1033 shape, the empty slit remained present for another 80 ms. A red fixation target was present  
1034 during the entire trial on top of the stimuli. Its continuous fixation resulted in a fluid reward  
1035 at the end of the trial. The shape fragments, presented in the first 4 frames (75 Hz frame  
1036 rate) of a vertical slit condition in which the shape moved leftwards (RL condition) are shown  
1037 to the right. Below it, we show shape fragments of the corresponding “random” condition. In  
1038 that condition, each of 9 shape fragments was presented for 6 frames (80 ms), except the  
1039 first and the last one which was presented for 2 frames. Below the slit view examples of the  
1040 random condition, the spatially concatenated shape parts that were presented during this  
1041 random condition are shown with the first presented fragment to the left. **D.** Delayed  
1042 matching to sample (DMS) test. After 300 ms of fixation of the red fixation target, on top of a  
1043 square noise background, the sample stimulus was presented. This could be a slit-viewing  
1044 movie with identical parameters as in the PF task: 480 ms of empty slit presentation,  
1045 followed by a display of the shape fragments for 773 ms and the empty slit for 80 ms. An  
1046 actual shape fragment of a single frame is shown for the horizontal and vertical slit

1047 presentation. The sample stimulus could also be a presentation of the whole shape (not  
1048 illustrated). Then, the background noise pattern without slit was presented for 53 ms after  
1049 which two shapes were presented simultaneously. The shapes were presented above and  
1050 below the fixation target. One of the two shapes corresponded to the one shown as sample  
1051 stimulus and the monkey was required to make a saccade to the stimulus that matched the  
1052 sample stimulus. The match and non-match stimuli stayed on the screen until the monkey  
1053 made a saccade to one of the stimuli or for maximally 4 s. Only correct responses were  
1054 rewarded with a fluid reward.

1055

1056 **Figure 2.** Example neuron recorded in the slit-viewing test during passive fixation.  
1057 Responses to static whole-shape presentations are shown for the best, medium, and worst  
1058 shape in the left column PSTHs. The full red line indicates shape onset. The 4 right columns  
1059 show the responses in the slit-viewing conditions. From left to right: rightward motion,  
1060 leftward motion, upward motion, and downward motion of the shape behind the slit. The  
1061 first, third, and fifth row show the PSTHs for the original slit-viewing conditions for the best,  
1062 medium, and worst shape, respectively. The second, fourth, and sixth rows (blue shading)  
1063 correspond to the random slit-viewing conditions. The full red line and 0 indicate the onset  
1064 of the slit, while the stippled red lines show the period in which the shape fragments were  
1065 visible through the slit. Bin width is 50 ms. This neuron was recorded in monkey MG.

1066

1067 **Figure 3.** Three example neurons recorded in the slit-viewing test during passive  
1068 fixation. The neurons were recorded in monkey MT (A), MB (B), and MG (C). Same  
1069 conventions as in Figure 2.

1070

1071

1072

1073 **Figure 4.** Population responses in the slit-viewing test during passive fixation. **A.**1074 Population PSTHs of the response to the best (blue), medium (green), and worst (red) static,  
1075 whole-shape presentations are shown in the left column. The responses in the slit-viewing  
1076 conditions are shown in the other 4 columns, with motion directions being rightward (LR),  
1077 leftward (RL), upward (DU) and downward (UD) from left to right (as in Figures 2 and 3). The  
1078 responses to the original and random slit-viewing conditions are shown in bold and light blue  
1079 color, respectively. Bands indicate the standard error of the mean. We normalized the  
1080 responses of each neuron by its maximum firing rate (bin width 10 ms) across all conditions.1081 Normalized responses of the 185 neurons of the three monkeys were averaged. Note the  
1082 different y-axes for the static and slit-viewing conditions. Other conventions as in Figures 2  
1083 and 3. **B.** Mean firing rate of the 185 neurons (3 monkeys) in each of the slit-viewing  
1084 conditions, for the best, medium, and worst shape. Responses are shown per motion  
1085 direction and for the original (o) and random (r) slit-viewing condition. Error bars indicate1086 the standard error of the mean. The mean firing rate (in spikes/s) was computed using an  
1087 analysis window of 800 ms, starting 50 ms after the shape became visible through the slit. **C.**  
1088 Mean firing rate of the 126 neurons (3 monkeys) that were tested during passive fixation of  
1089 the original (o), jumping (j), and random (r) slit-viewing conditions. Responses were sorted  
1090 per motion direction, shape, and viewing condition. Same conventions as in **B.**

1091

1092 **Figure 5.** Decoding of shape identity from the responses of the recorded neurons  
1093 ( $n=185$  neurons with 10 trials per condition). **A.** Classification accuracy (% correct) when the  
1094 classifier was trained with the responses to static shapes using the long 800 ms analysis

1095 window. Testing was performed for the static (S) and the four slit-viewing motion directions  
1096 (LR, RL, UD, DU). Chance performance (33%) is indicated by the horizontal stippled line. **B**.  
1097 Classification accuracies when training shape identification using the responses of the LR slit-  
1098 viewing condition and testing each of the 5 conditions. Error bars indicate standard deviation  
1099 of the classification scores across resamplings. All classification scores were significant  
1100 (permutation test with shuffled condition labels), except those indicated by n.s. **C**.  
1101 Classification accuracies when training shape identification using the responses of the RL  
1102 condition. Same conventions as in **A** and **B**. **D**. Classification accuracies when training shape  
1103 identification using the responses of the UD condition. Same conventions as in **A** and **B**. **E**.  
1104 Classification accuracies when training shape identification using the responses of the DU  
1105 condition. Same conventions as in **A** and **B**. **F**. Classification scores when training and testing  
1106 were done using bins of 100 ms duration. Training and testing were performed for all  
1107 combinations of bins and conditions. Each panel shows the classification scores with the  
1108 trained bin along the vertical axis and the tested bin along the horizontal axis. Panels of a  
1109 row have the same trained condition, while the columns correspond to the tested  
1110 conditions. The white vertical and horizontal line in each panel marks the onset of the first  
1111 visible shape fragment. The red stippled square outlines indicate panels where trained and  
1112 tested conditions had the same slit orientation. Percent correct classifications are presented  
1113 with a color code. The same population of neurons as in **A-E**.

1114

1115 **Figure 6.** Responses in slit-viewing and snapshot tests compared. **A**. Example neuron,  
1116 recorded in monkey MG. PSTHs (bin width = 10 ms) show the responses of the neuron in the  
1117 LR (left PSTH) and RL (right) original slit-viewing conditions. Conventions as in Figure 2. The  
1118 panels below the PSTHs show the responses to the shape segments of individual frames

1119 using a color code, with hot and cold colors indicating a strong and weak response,  
1120 respectively. The responses in these “shape response plots” were computed using a bin of  
1121 75 ms duration and a delay of 70 ms relative to shape fragment onset. **B.** The first row shows  
1122 the responses in the snapshot test for the same neuron as in **A**. Stimulus onset is indicated  
1123 by the full red line. The shape response plot (second row) shows the normalized firing rate to  
1124 the 11 shape fragments in color code, using the same convention as in **A**. The middle lower  
1125 panel (binned activity) shows the mean normalized firing rate for 11 bins of the shape  
1126 response plots for each of the two motion directions presented in **A**, and the snapshot  
1127 condition. Note that we reversed the order of the shape segments of the RL slit-viewing  
1128 condition (yellow) to have corresponding shape parts in the different vertical slit conditions  
1129 (LR slit-viewing (red) and snapshot (blue) test conditions). The responses were normalized by  
1130 the maximum response for each condition separately and Pearson correlation coefficient ( $r$ )  
1131 was calculated between the binned responses of the slit-viewing conditions and the  
1132 snapshot condition. **C.** Distribution of the Pearson correlation coefficient between the  
1133 binned responses to the shape fragments during slit-viewing and in the snapshot test. Data  
1134 of the two monkeys are shown separately and the medians are indicated by the colored  
1135 triangles. The number of observations ( $n$ ) corresponds to slit-viewing direction - shape  
1136 combinations. **D.** Shape response plots for two opponent motion directions of the slit-  
1137 viewing and corresponding snapshot tests for other example neurons (rows) recorded in  
1138 monkeys MG and MT. The same conventions as in **B**. Examples of vertical (upper two rows)  
1139 and horizontal (bottom two rows) slit orientations are shown for each monkey.  
1140  
1141  
1142

1143       **Figure 7.** Behavioral results in the DMS task and eye movements. **A.** Behavioral  
1144 accuracy (% correct) in the DMS task for different combinations of sample and test stimulus  
1145 conditions. S: static whole-shape sample presentation; Slit: slit-view sample stimulus; N: new  
1146 stimulus; O: old stimulus, X/Y: X is match and Y non-match stimulus. Thus, the condition “Slit  
1147 N/O” corresponds to trials in which the sample stimulus was a slit-view presentation of a  
1148 new stimulus, the match was a new shape, and the non-match an old shape. Error bars  
1149 indicate 95% confidence intervals. The horizontal red line corresponds to the chance level  
1150 (50 % correct). **B.** Mean horizontal and vertical eye position and 95% confidence intervals  
1151 during slit-viewing in the passive fixation (PF) task for each of the 4 motion directions,  
1152 obtained during the recordings in monkey MT. Before averaging, we subtracted for each trial  
1153 the mean eye position in a 20 to 0 ms period before motion onset from the eye positions  
1154 after motion onset. For each test session of a neuron, the baseline-subtracted eye positions  
1155 were first averaged across the trials in which the same shape was presented and then  
1156 averaged across condition-neuron combinations. The number of observations (n)  
1157 corresponds to the number of condition-neuron combinations. **C.** Same as **B.**, but data of  
1158 monkey MB in the PF task. **D.** Eye position data for monkey MG obtained during the  
1159 recordings in the PF task before DMS task training (upper panels), in the DMS task (middle  
1160 panels), and the PF task after training of the DMS task (post-DMS; bottom panels). Same  
1161 conventions as in **B.**

1162

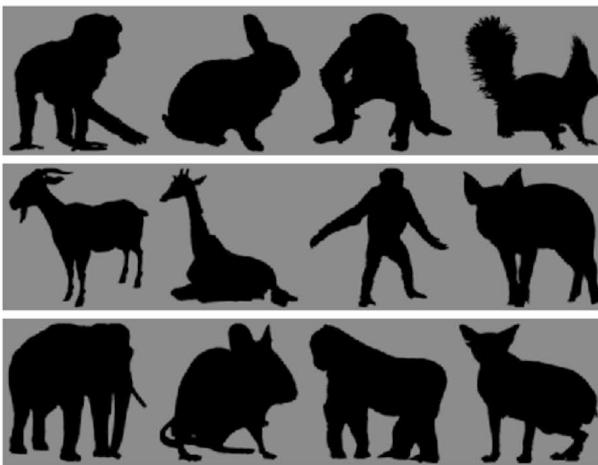
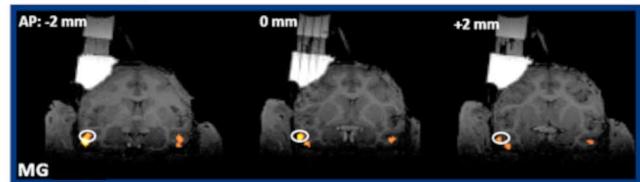
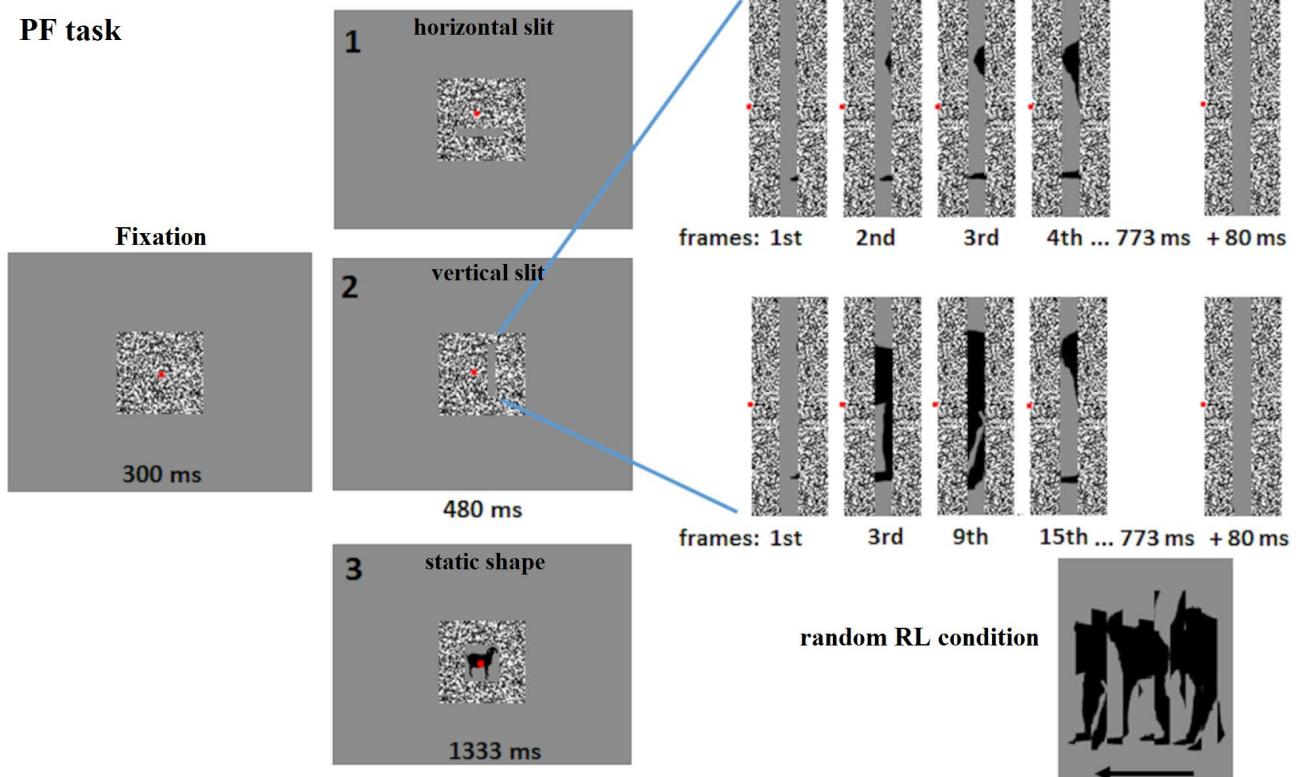
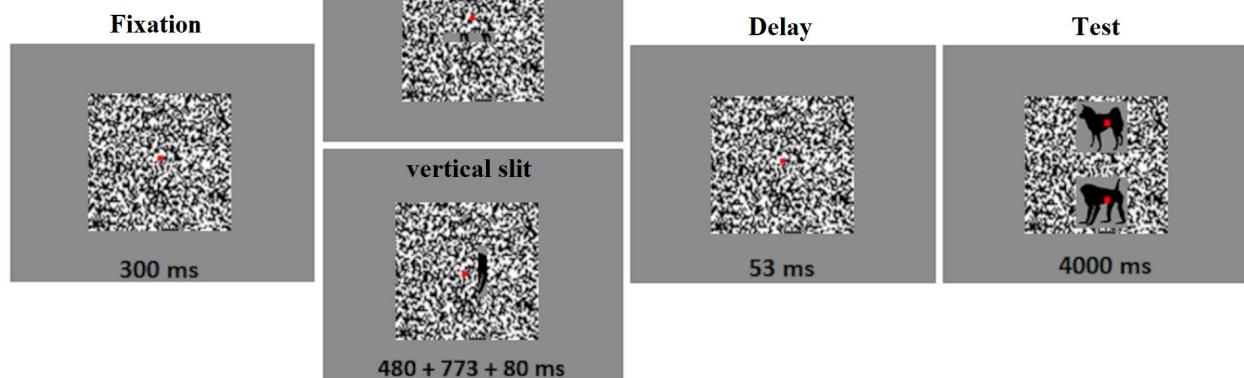
1163       **Figure 8.** Distributions of response properties in the PF and DMS tasks during slit-  
1164 viewing. **A.** Distribution of the same-axis Direction Index for the neurons recorded in the PF  
1165 task for each of the 3 monkeys. Triangles indicate medians. **B.** Distribution of the same-axis  
1166 Direction Index for neurons recorded in the PF task before training, during the DMS task, and

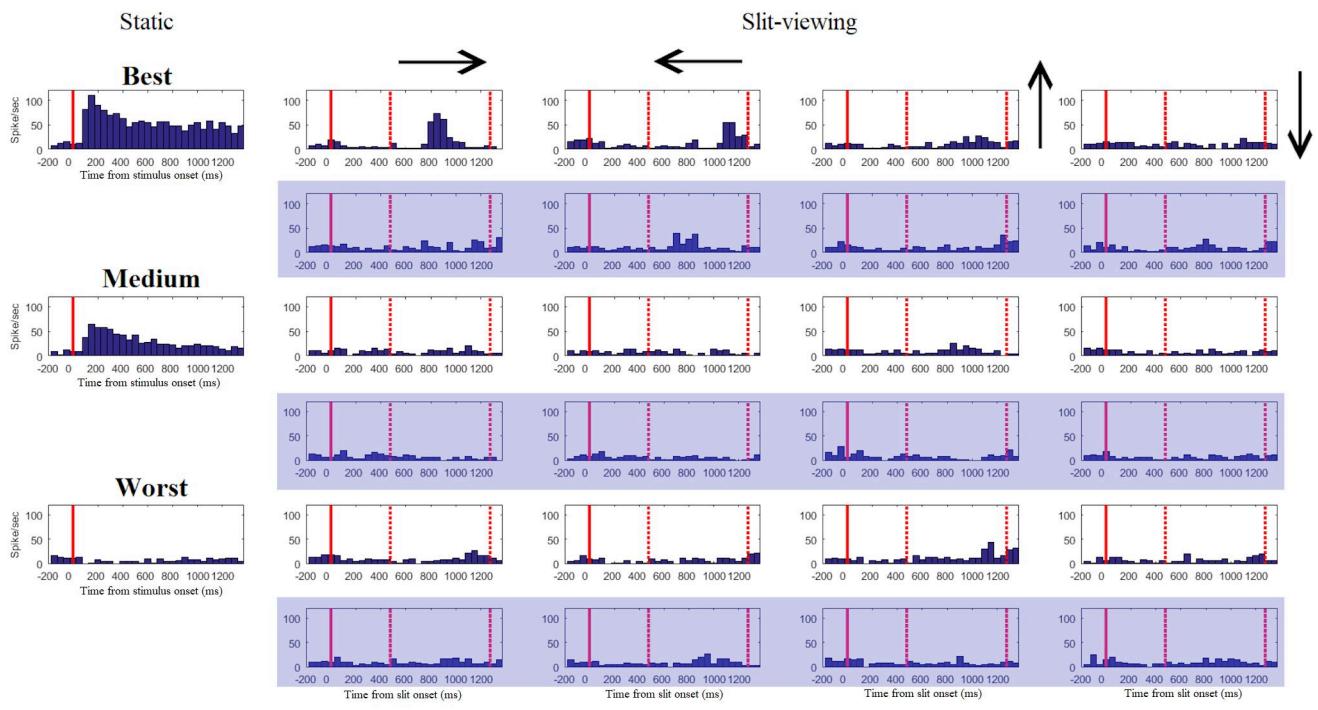
1167 in the post-DMS task. Data of monkey MG. **C.** Distribution of the Direction Index for  
1168 orthogonal axes for the neurons recorded in the PF task for each of the 3 monkeys. Each  
1169 neuron contributed two Direction Index values, one for each of the two directions along the  
1170 orthogonal axis. **D.** Distribution of the Direction Index for orthogonal axes for neurons  
1171 recorded in the PF task before training, during the DMS task, and in the post-DMS task. Data  
1172 of monkey MG. **E.** Distribution of the peak duration metric for the neurons recorded in the  
1173 PF task for each of the 3 monkeys. **F.** Distribution of the peak duration metric for the  
1174 neurons recorded in the PF task before DMS training, during the DMS task, and post-DMS  
1175 training in monkey MG. **G.** Distribution of the Separability Index for the neurons recorded in  
1176 the PF task before DMS training, during the DMS task and post-DMS training in monkey MG.

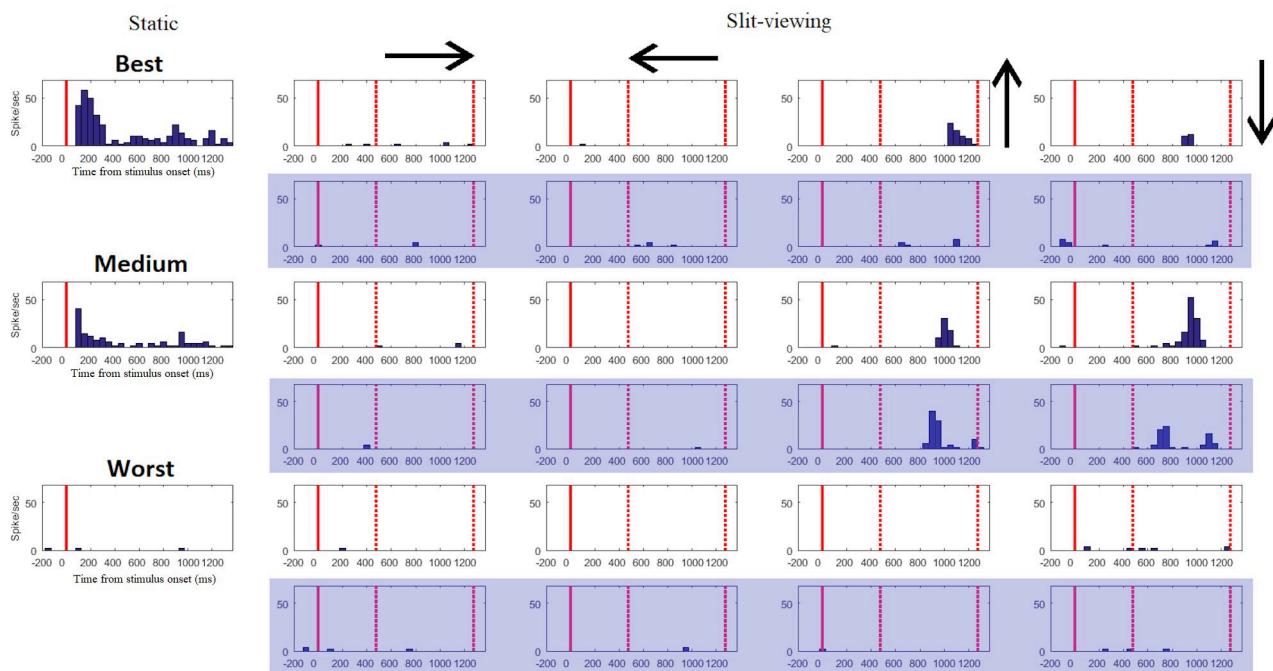
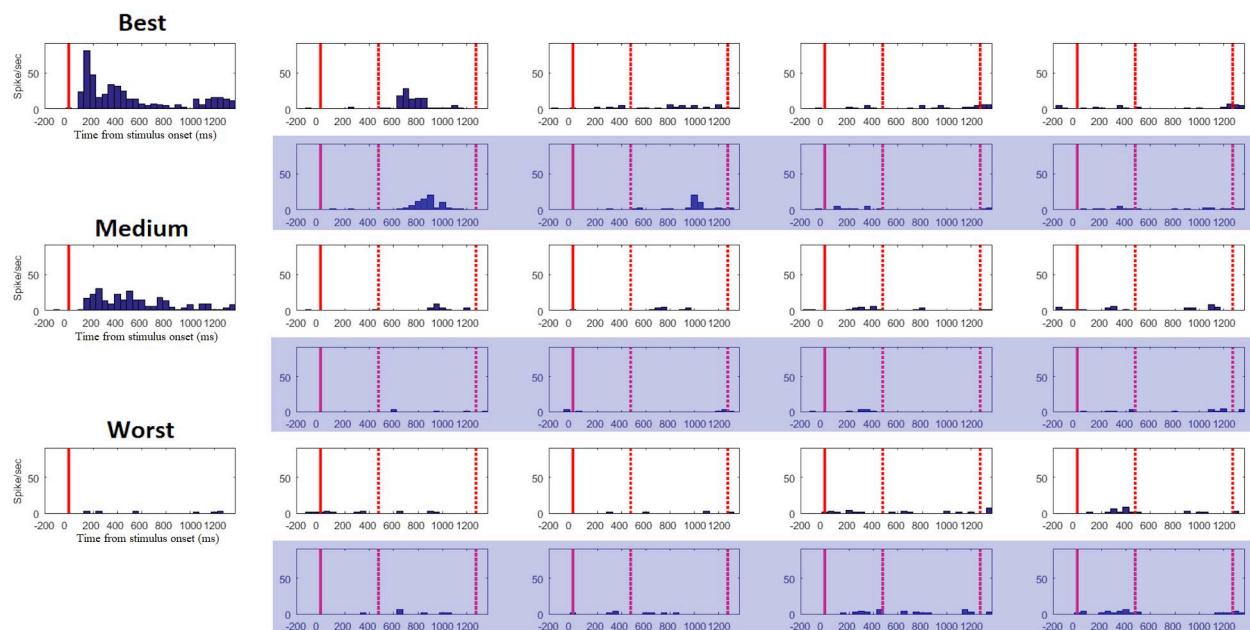
1177

1178

1179

**A Example stimuli****B****Recording location****C****PF task****D****DMS task**



**A****B****C**