

1 **Mice and primates use distinct strategies for visual segmentation**

2
3 **Francisco J. Luongo^{1*}, Lu Liu^{1*}, Chun Lum Andy Ho², Janis K. Hesse^{1,3}, Joseph B. Wekselblatt¹,**
4 **Francesco Lanfranchi^{1,3}, Daniel Huber², and Doris Y. Tsao^{1,3,4**}, †**

- 5
6 1. Division of Biology and Biological Engineering, Caltech, Pasadena CA 91125
7 2. University of Geneva, Department of Basic Neurosciences, Rue Michel Servet 1, Geneva 1211,
8 Switzerland
9 3. Computation and Neural Systems, Caltech, Pasadena CA 91125
10 4. Howard Hughes Medical Institute, Pasadena, CA 91125

11 * Co-first author

12
13 ****Corresponding author:**

14 Doris Y. Tsao
15 California Institute of Technology
16 Division of Biology
17 MC 114-96 Pasadena, CA 91125
18 USA

19
20 Phone: 626-395-1702 Fax: 626-395-8826

21 Email: dortsao@caltech.edu

22
23 †Lead Contact

24
25 **Keywords**

26 mouse vision, segmentation, motion perception, object localization

27

28 **Abstract**

29 The rodent visual system has attracted great interest in recent years due to its experimental tractability, but
30 the fundamental mechanisms used by the mouse to represent the visual world remain unclear. In the
31 primate, researchers have argued from both behavioral and neural evidence that a key step in visual
32 representation is “figure-ground segmentation,” the delineation of figures as distinct from backgrounds [1-
33 4]. To determine if mice also show behavioral and neural signatures of figure-ground segmentation, we
34 trained mice on a figure-ground segmentation task where figures were defined by gratings and naturalistic
35 textures moving counterphase to the background. Unlike primates, mice were severely limited in their ability
36 to segment figure from ground using the opponent motion cue, with segmentation behavior strongly
37 dependent on the specific carrier pattern. Remarkably, when mice were forced to localize naturalistic
38 patterns defined by opponent motion, they adopted a strategy of brute force memorization of texture
39 patterns. In contrast, primates, including humans, macaques, and mouse lemurs, could readily segment
40 figures independent of carrier pattern using the opponent motion cue. Consistent with mouse behavior,
41 neural responses to the same stimuli recorded in mouse visual areas V1, RL, and LM also did not support
42 texture-invariant segmentation of figures using opponent motion. Modeling revealed that the texture
43 dependence of both the mouse’s behavior and neural responses could be explained by a feedforward
44 neural network lacking explicit segmentation capabilities. These findings reveal a fundamental limitation in
45 the ability of mice to segment visual objects compared to primates.

46 **Introduction**

47 Primates rely primarily on vision to meaningfully interact with objects in the world. Mice, in contrast, rely far
48 less on their visual system, though they do use visual cues for important behaviors such as hunting,
49 evasion, and navigation [5-9]. The field of mouse vision has attracted great excitement in recent years due
50 to the wealth of tools available for mouse circuit dissection, with many groups adopting the mouse as a
51 model for visual perception [10] and visually-guided decision making [11, 12]. Yet the fundamental ethology
52 of mouse vision remains poorly understood. What does the mouse perceive as a visual object?
53

54 While work has shown that visual responses in mouse visual cortex share low-level organizing principles
55 with those of primate visual cortex, including temporal/spatial frequency tuning [13], orientation selectivity
56 [14], and contextual surround effects [15, 16], it remains unclear to what extent the two species share more
57 abstract representations of visual objects and scenes.
58

59 In particular, it is unclear whether mice explicitly segment visual scenes into discrete surfaces.
60 Segmentation refers to the identification of borders of each object in a visual scene and assignment of
61 discrete labels to pixels corresponding to each object. In primates, segmentation is a key step in visual
62 processing following early feature extraction [2, 17-20]. For example, in the famous “face-vase” illusion,
63 human viewers inexorably segment the scene as a face or a vase, with bistable dynamics. A large body of
64 psychophysics suggests that the primate visual system performs segmentation by generating a *surface*
65 *representation*, an assignment of each retinal pixel to a distinct contiguous surface situated in 3D space
66 [18, 21] (**Fig. 1a**).
67

68 How could the brain solve visual segmentation? The key visual cue signaling a surface border is a
69 *discontinuity*, an abrupt change in features at the surface border. For example, there is often a change in
70 luminance, orientation, or texture at a surface border. However, this need not be the case: changes in
71 luminance, orientation, and texture can also occur within interior regions of a surface (**Fig. 1b**). Conversely,
72 object borders can exist without any change in luminance, orientation, or texture—a fact exploited by
73 animals that use camouflage [22]. Thus a key challenge of intermediate vision is to identify true object
74 borders using ambiguous local cues. Aiding this goal, there is one cue that is unambiguous: *accretion-*
75 *deletion*, the appearance or disappearance of pixels forming the background surface due to motion (or
76 binocular disparity) of the foreground surface (**Fig. 1b**). Gibson identified accretion-deletion as the *single*
77 *most important cue to surface organization* because it is unambiguous, invariant to texture, and locally
78 available [23]. Psychophysical experiments in humans demonstrate that accretion-deletion alone is able to
79 evoke a vivid percept of an object border [24]. Furthermore, a recent computational theory of surface
80 representation shows how surface segmentation can be computed using local accretion-deletion cues in a
81 simple way without requiring learning, top-down feedback, or object recognition [21]. Moreover, this new
82 theory shows how such local accretion-deletion cues can be used not only to solve segmentation, but also
83

84 to solve *invariant tracking of objects* (**Fig. 1c**), widely considered one of the hardest problems in vision [25,
85 26]. To summarize, while there are multiple cues to segmentation, accretion-deletion holds a special status
86 for the following reasons: (i) it is unambiguous, occurring only at true object borders and never at internal
87 texture borders; (ii) it is cue-invariant; (iii) it is especially computationally powerful, supporting not only
88 segmentation but also invariant tracking.

89
90 A variety of neural correlates of segmentation have been found in the primate brain. Neurons in primate
91 V1, V2, and V4 modulate their firing according to whether a stimulus is part of the foreground or background
92 [2-4] (**Fig. 1d**). Complementing figure-ground signaling, a population of neurons have been found in
93 macaque areas V2 and V4 that explicitly signal object borders. These “border-ownership” cells respond
94 selectively to figure edges and are moreover modulated by the side of the figure relative to the edge [19]
95 (**Fig. 1e**).

96
97 It remains unclear whether figure-ground and border-ownership cells provide an explicit, cue-invariant
98 solution to the segmentation problem, for both conceptual and experimental reasons. Conceptually, natural
99 vision scenes typically contain multiple objects, necessitating multiple distinct neural labels. However,
100 figure-ground cells seem capable of unambiguously labeling only a single figure via increased firing rate;
101 additional processing of figure-ground and border-ownership cell output would be necessary to generate
102 distinct labels for each object. Experimentally, classic studies of figure-ground segmentation in the primate
103 used figures composed of oriented lines on a background of orthogonally-oriented lines [2]. Such
104 modulation could in theory have arisen as a result of lower-level cues (e.g., neurons with receptive fields
105 near the edge signaling low-level orientation contrast) [27]. Hesse and Tsao investigated the consistency
106 in the sign of selectivity of border-ownership cells across different visual conditions. While 55% of V2/V3
107 cells met the criteria for border ownership selectivity when assayed with a luminance-defined square, not a
108 single cell showed fully consistent selectivity when assayed with figures defined by other types of cues [28].
109 Overall, these experimental findings raise the possibility that segmentation is solved by a population code
110 rather than explicit cue-invariant figure and edge detectors. Despite these uncertainties about neural
111 mechanism, we underscore that behavioral evidence unequivocally demonstrates that primates possess a
112 mechanism for explicit, cue-invariant segmentation exploiting accretion-deletion [24].

113 Behaviorally, mice are capable of texture-based segmentation, in which figure and background are defined
114 by grating patterns and have different orientation or phase [29, 30]. Consistent with this behavioral
115 capability, cells in mouse V1 show iso-orientation surround suppression [15, 16] and have been reported
116 to be modulated by figure versus ground [29-31]. However, all of these studies have used texture-based
117 cues, which are fundamentally ambiguous for solving segmentation (**Fig. 1b**). *Thus it remains an open*
118 *question whether mice are capable of explicit object segmentation, or simply of texture segregation.*

119
120 Here, we took advantage of the ability to record from large numbers of neurons across the mouse cortical
121 visual hierarchy to look for behavioral and neural correlates of visual segmentation in the mouse. We
122 discovered a surprising difference between mouse and human segmentation behavior, which led us to
123 systematically investigate segmentation behavior in three additional species: the macaque, mouse lemur,
124 and treeshrew. We found that the mouse, like the treeshrew and unlike the two primate species, is
125 behaviorally incapable of texture-invariant segmentation. In fact, mice tasked to localize objects with
126 naturalistic textures adopted a strategy of brute force memorization—a cognitively impressive feat.
127 Furthermore, we found no evidence for single neurons in mouse visual cortex modulated by figure/ground
128 or border ownership in a texture-invariant manner. For patterns containing orientation or phase contrast
129 between figure and background, we could decode figure location from population neural recordings, with
130 best decoding in putative ventral stream area LM, followed by RL and V1, but we could not decode figure
131 location for figures with naturalistic texture. A simple feedforward neural network could account for the
132 observed dependence of mouse behavior and neural responses on carrier pattern. Taken together, these
133 findings reveal a fundamental difference between primate and mouse mechanisms for object segmentation,
134 with the mouse relying much more on texture statistics than the primate. The findings have broad
135 implications for use of the mouse as a model for visual perception.

136 137 **Results**

138

139 **Mice fail to segment objects defined purely by opponent motion**

140
141 We set out to clarify (i) whether mice are capable of invariantly segmenting figure from ground and (ii)
142 whether there exist texture-invariant segmentation-related signals in the mouse brain, as reported in the
143 macaque brain [2, 32]. To address the mouse's ability to segment objects, we designed a two-alternative
144 forced choice task in which mice reported the side of a touch screen that contained a figure for a water
145 reward (**Fig. 2a**). We tested mouse segmentation behavior using three classes of stimuli: (i) "Cross" stimuli,
146 in which the figure consisted of a grating, and the ground consisted of an orthogonal grating; (ii) "Iso" stimuli,
147 in which the figure consisted of a grating, and the ground consisted of a grating at the same orientation, but
148 offset in phase; (iii) Naturalistic ("Nat") stimuli, in which both the figure and the ground consisted of 1/f noise
149 patterns (**Fig. 2b, Extended Data Movie 1**). In each case, figure and ground moved in counterphase
150 providing a differential motion cue with accretion-deletion; this motion cue was essential for defining the
151 figure in the Nat condition. The logic of including these three conditions was as follows: (i) the Cross
152 condition has been used previously in multiple studies of figure ground segmentation [2, 3] and extra-
153 classical receptive field modulation [15]; (ii) the Iso condition constitutes a slightly more challenging figure-
154 ground segmentation problem due to lack of orientation contrast; nevertheless, the figure can be readily
155 segmented in static images using the phase difference between figure and ground; (iii) the Nat condition
156 allowed us to disambiguate true figure-ground signals from low-level orientation or phase contrast signals.

157
158 We first trained mice on four different patterns (two orientations/textures x two sides) for each of the three
159 stimulus conditions (Cross, Iso, Nat, **Fig. 2b**). Each session consisted of a single condition (see Methods).
160 The learning curves for the three stimulus conditions were very different (**Fig. 2c**). Mice quickly learned the
161 Cross task, reaching 88% performance after 7 days. They were slightly slower to learn the Iso task, reaching
162 77% performance after 9 days. However, they struggled to effectively learn the Nat task, reaching only
163 around 71% performance after 13 days. We next tested two macaque monkeys on the same task. The
164 monkeys performed at >90% for all three conditions within the first session (**Fig. 2d**). This is consistent with
165 Gibson's original observation that accretion-deletion provides a strong cue to object borders in humans
166 [23], as well as previous studies showing that figures defined by differential motion cues can be readily
167 detected by macaque monkeys [2]. Thus, there was a clear difference between the segmentation
168 capabilities of the mouse and primate.

169
170 We next wondered whether through more gradual shaping, the mice could learn the Nat task. We trained
171 the mice in a series of stages across 26 training sessions over which the stimulus morphed from the Cross
172 to the Nat condition (**Fig. 2e**). For each stage, mice would reach good performance (> 80%), followed by a
173 dramatic drop when a new, more difficult stage was introduced. By the end of 26 training sessions, three
174 out of four mice successfully learned to detect the square in the full Nat condition (**Fig. 2e**, "100%"). Thus
175 it appeared that through this gradual shaping, mice had acquired the ability to segment figure from ground
176 using opponent motion.

177
178 To confirm this, we next tested the mice on seven new textures. To our surprise, the mice performed near
179 chance on these new textures (**Fig. 2f**, mean performance across three mice that had learned the task:
180 60%; test for significant difference between performance on new textures and performance on last day of
181 noise shaping: $p < 0.01$ for each mouse, Chi-square test). This lack of ability to generalize suggests that the
182 mice had not in fact learned to segment figure from ground using opponent motion.

183
184 How then were they able to perform the Nat task on the trained patterns? We hypothesized that the animals
185 had simply learned to memorize the mapping between the noise patterns in the Nat condition and the
186 appropriate motor response, in effect using a lookup table from four patterns to two actions instead of relying
187 upon visual perception of a figure. If this was the case, we reasoned that we should be able to remove
188 motion from the stimulus and the animals should still perform well. Astonishingly, this turned out to be the
189 case: mice displayed no change in performance upon removal of motion in the stimulus, which completely
190 removed any way of inferring a figure (**Fig. 2g**, mean performance across three mice that had learned the
191 task: 87%; test for significant difference between performance on static textures and performance on last
192 day of noise shaping: $p > 0.01$ for each mouse, Chi-square test).

193

194 We next tested all three conditions (Cross/Iso/Nat) in a separate cohort of mice to further examine whether
195 the animals were indeed discarding the opponent motion cue. We tested them on a static condition of the
196 task after training on the motion task. As before, these mice showed no drop in behavioral performance for
197 any of the three conditions (**Fig. 2h**, Cross: $p=0.67$, Iso: $p=0.02$ (increasing), Nat: $p=0.26$), confirming that
198 the animals were not using the motion cue for figure detection. While this was not surprising for the Cross
199 and Iso cases, as the single static frame still had strong edge contrast due to orientation/phase differences
200 and thus contained a clear figure that could be detected, it was surprising for the Nat condition which had
201 minimal cues for the figure when static. Thus this experiment further confirmed that mice did not use the
202 opponent motion cue to perform the segmentation task.

203
204 For comparison, we performed the same test on two monkeys. Their performance showed a very different
205 pattern. Like mice, monkeys did not display a drop in performance in the Cross and Iso conditions. For the
206 Nat condition, however, monkeys showed a dramatic drop in performance when motion cues were removed
207 from the stimulus (**Fig. 2h**, Fisher Exact test, p -vals, Monkey 1: Cross 1.0, Iso: 0.407, Nat: $9.7e-14$; Monkey
208 2: Cross 0.77, Iso: 0.75, Nat: $2.21e-19$). This experiment reveals that monkeys and mice used
209 fundamentally different strategies to solve the Nat condition: monkeys used the opponent motion cue,
210 resulting in their dramatic drop in performance upon removal of motion, while mice used a learned lookup
211 table mapping patterns to responses.

212
213 Given the inability of mice to generalize to new textures for the Nat condition (**Fig. 2f**), we wondered whether
214 the same would hold true for the Cross and Iso conditions. We next trained 8 new animals on the Cross
215 and Iso tasks with 4 patterns (2 orientations \times 2 positions), and then tested their ability to generalize to 10
216 new patterns from the same class (see Methods). We found that mice were able to generalize well for the
217 Cross condition (**Fig. 2i1**, left, mean performance drop = 2.47%), and moderately well for the Iso condition
218 (mean performance drop = 6.55%). However, for the Nat condition, performance dropped almost to chance
219 (mean performance drop = 20.61%; $p=0.0011$, ranksum test), consistent with the result of our previous
220 experiment (**Fig. 2f**). One concern may be that the mice simply could not see the motion cue. To control for
221 this, we trained the same four mice to perform the Nat condition in a modified situation in which the
222 background was static. The animals learned this modified Nat task much more quickly, indicating that they
223 could see the motion cue (**Extended Data Fig. 1a**).

224
225 Since mice showed their best generalization performance for cross-oriented gratings, this suggested that
226 orientation contrast is a key feature used by mice to solve the task of localizing the figure. This in turn
227 suggested that we might improve the animal's performance on random textures by introducing an element
228 of orthogonal orientation. To test this, we generated two new sets of figure-ground stimuli starting from
229 random textures: (i) "Iso-tex" stimuli, in which a square was cropped from the texture and placed in either
230 the left or right position, and (ii) "Cross-tex" stimuli, in which the same square was rotated 90° , increasing
231 orientation contrast (**Extended Data Fig. 2a**); for both sets of stimuli, opponent motion between figure and
232 ground was added. To compare the generalization ability of mice on these two classes of stimuli, we first
233 measured baseline performance on Iso-tex and Cross-tex stimuli drawn from 7 random textures. We then
234 trained mice Iso-tex and Cross-tex stimuli drawn from a different set of 30 textures. Finally, we re-measured
235 performance on the original set of 7 textures (**Extended Data Fig. 2b**). While there was no difference in
236 baseline performance between the two conditions, a significant difference emerged during training
237 (**Extended Data Fig. 2c-f**). Animals trained on the Iso-tex condition achieved a mean performance of 57%
238 after 14 days of training (**Extended Data Fig. 2c, d, g**), whereas animals trained on the Cross-tex condition
239 achieved 67% correct after 14 days (**Extended Data Fig. 2e, f, g**), indicating that a strong orthogonal
240 component could aid the mice in performing the task. However, despite above chance performance on the
241 bank of 30 random textures, just as before, mice were largely unable to utilize any information about the
242 motion cue, as demonstrated by their drop back to initial performance for the original bank of 7 textures
243 (**Extended Data Fig. 2d, f, h**). Overall, our behavioral results suggest that mice adopt a strategy for object
244 localization that relies heavily on orientation contrast and phase differences between figure and ground and
245 is blind to opponent motion cues.

246 247 **Comparing segmentation behavior in mouse, macaque, treeshrew, and mouse lemur**

248 The striking difference between mouse and macaque segmentation behavior inspired us to run the
249 generalization test of **Fig. 2i1** on two macaque monkeys (**Fig. 2i4**). The macaques showed a very different

250 behavioral pattern compared to the mice: they were able to generalize to unseen patterns for all 3
251 conditions, indicating that they were capable of performing segmentation using the opponent motion cue,
252 and had not simply memorized the texture pattern in the Nat condition like the mice.

253
254 This clear difference between the behavioral strategies for visual segmentation used by mice versus
255 macaques further inspired us to perform the same pattern generalization test (i.e., train on one set of
256 patterns/orientations, test on a different set of unseen patterns/orientations) in two additional species: (i) a
257 second mammalian species of the order scandentia (*Tupaia belangeri*; treeshrew), and (ii) a second primate
258 species (*Microcebus murinus*; mouse lemur). The treeshrews performed similarly to mice, displaying
259 generalization for the Cross and Iso conditions but not the Nat condition (**Fig. 2i2**). In contrast, and similarly
260 to macaques, the mouse lemurs were readily able to generalize for all three conditions (**Fig. 2i3**), implying
261 that they, like the macaques (**Fig. 2i4**), were able to perform visual segmentation using the opponent motion
262 cue. Training curves for all four species on this task are shown in **Extended Data Fig. 1b & c**. Taken
263 together, these results provide strong evidence that primates including mouse lemurs, macaques, and
264 humans all use a visual segmentation strategy exploiting opponent motion cues, in contrast to mice and
265 treeshrews, which rely on texture cues to perform visual segmentation and are incapable of using opponent
266 motion cues.

267 268 **Absence of texture-invariant figure signals in mouse visual cortex**

269
270 Given the evident inability of mice to perform texture-invariant visual segmentation, a natural question
271 arises: what segmentation-related signals are available in mouse visual cortex to decode the location and
272 boundary of an object? To address this, we recorded responses in mouse visual cortex to figure-ground
273 stimuli defined by both texture and opponent motion using (i) electrophysiology with a 64-channel silicon
274 probe, and (ii) 2-photon calcium imaging. We compared responses across three distinct mouse visual
275 areas: primary visual cortex (V1), a putative ventral stream area (LM), and a putative dorsal stream area
276 (RL) [33].

277
278 We first localized visual areas using wide-field imaging in GCAMP6s transgenic animals and used
279 vasculature maps to guide subsequent two photon and electrophysiology experiments (**Fig. 3a, b**) (see
280 Methods). We then measured the receptive field centers of neurons using either a sparse noise stimulus or
281 a spatially isolated flashing Gabor of varying orientations. Imaging and electrophysiology data were
282 generally consistent. For most analyses below (with the exception of **Fig. 3f, g** and **Fig. 6d, f**), we present
283 electrophysiology data, as it had better temporal resolution and gave better receptive field estimates (due
284 to absence of neuropil activity leading to blurring of receptive fields). The latter was critical as the analyses
285 depended on accuracy of receptive field estimates for single cells.

286
287 To visualize a neuron's response to figure, ground, and borders, we computed a "figure map" consisting of
288 the neuron's mean response to a figure centered at each of 128 (16 x 8) positions across the visual field
289 (**Fig. 3c, Extended Data Movie 2**). The square figure appeared for 250 ms at each position. The stimulus
290 closely mimicked that used for behavioral tests; in particular, the square moved in counterphase to the
291 background. This stimulus enabled us to measure responses of each neuron to figure, ground, and borders,
292 as on any given trial a particular location contained figure, ground, or borders. For example, the figure map
293 of an ideal figure cell would reveal a square corresponding to the figure (**Fig. 3d, left**), that of a border cell
294 would reveal stripes corresponding to the figure borders matching the orientation of the cell (**Fig. 3d,**
295 **middle**), and that of an ON-cell would reveal phase-dependent responses to the figure (**Fig. 3d, right**). As
296 these model units illustrate, the figure map is a function of a cell's receptive field location, low-level stimulus
297 preferences (e.g., orientation selectivity, contrast polarity selectivity), and high-level stimulus preferences
298 (figure/ground selectivity, border selectivity). Thus the figure map yields a rich fingerprint of a cell's visual
299 selectivity.

300
301 **Fig. 3e1** shows the figure map for one example cell from V1 recorded with electrophysiology in the Cross
302 condition. Responses to a subset of four stimuli (**Fig. 3e2**) revealed a decreasing response as the figure
303 moved off the receptive field (**Fig. 3e1-e3**). We confirmed that these figure maps were highly stable using
304 two-photon imaging across multiple days. **Fig. 3f** shows figure response maps obtained from six example
305 cells across two different days. The mean correlation between maps from matched cell pairs across

306 different days was very high (N = 950 cell pairs, Pearson $r=0.5579$ matched vs Pearson $r=0.2054$
307 unmatched, KS test $p = 1e-163$, **Fig. 3g**).

308
309 To fully characterize responses of neurons to figure, ground, and borders, we obtained figure maps using
310 the same three conditions as those used earlier in behavior (Cross, Iso, Nat) (**Fig. 4a**). We presented two
311 orientations/textures for each of the three conditions. In V1, we often found cells that showed orthogonal
312 stripes for the two different cross patterns (**Fig. 4b**), as expected for an ON- or OFF-cell (**Fig. 3d**, right). We
313 failed to find any cells in V1, LM, or RL that showed consistent figure maps across the different conditions
314 (**Fig. 4a-d**). To quantify this across the population, we computed distributions of the mean Pearson
315 correlation between figure maps across all possible pairs from the six conditions: the values were centered
316 around 0 (V1: 0.032, LM: 0.034, RL: 0.015) (**Fig. 5a**). Within each condition, the mean Pearson correlation
317 between figure maps was also centered around 0 (V1:0.013, LM:0.008, RL:0.007) (**Fig. 5b**). This shows
318 that across the population, selectivity to figure location within individual neurons was strongly dependent
319 on the specific texture of the figure.

320
321 We next quantified selectivity for figure ground and border ownership, two of the hallmark segmentation-
322 related signals reported in the macaque visual system, across the V1, LM, and RL cell populations. We
323 only analyzed neurons that had significant receptive field fits (see Methods); furthermore, we confined our
324 analysis to neurons with receptive fields centered within four degrees of the monitor center, to ensure that
325 there were an adequate number of figure, ground, and left/right trials from which to compute the neuron's
326 modulation indices. For each of the three conditions, we defined a figure-ground modulation index as
327 $FGM = \frac{(R_{Fig} - R_{Back})}{(R_{Fig} + R_{Back})}$, where R_{Fig} is the mean response across the two patterns for the condition within the
328 figure zone, i.e., the 2×2 ($6^\circ \times 6^\circ$) grid of locations centered on the monitor (4 locations \times 10 trials = 40
329 figure trials) and R_{Back} is the mean response across the two patterns for the condition in the background
330 zone, i.e., the leftmost and rightmost column of locations (2×8 locations \times 10 trials = 160 background trials)
331 (**Fig. 5c**). We were extremely conservative in our selection of figure and ground locations to avoid any
332 mistakes in labeling due to uncertainties in receptive field location. Distributions of FGM indices were
333 centered on zero (Cross:-0.11, Iso:-0.04, Nat:-0.04) (**Fig. 5d**). To determine significance of FGM indices,
334 for each cell and condition, we estimated a bootstrapped p-value for the corresponding FGM value using
335 distributions of shuffled trials (see Methods); we found no neurons that showed significant figure modulation
336 across more than three conditions (**Fig. 5e**).

337
338 We quantified border ownership selectivity in a similar way. We defined a border ownership modulation
339 index as $BOM = \frac{(R_{Left\ border} - R_{Right\ border})}{(R_{Left\ border} + R_{Right\ border})}$, where $R_{Left\ border}$ is the mean response across the two patterns
340 for the condition within the left border zone (Column 4; 8 locations \times 10 trials = 80 left edge trials), and
341 $R_{Right\ border}$ is the mean response across the two patterns for the condition within the right border zone
342 (Column 12; 8 locations \times 10 trials = 80 right edge trials) (**Fig. 5f**). Distributions of BOM indices were
343 centered on zero (Cross:-0.02, Iso:0.01, Nat:-0.02) (**Fig. 5g**). We found no neurons that showed significant
344 border ownership modulation across more than three conditions (**Fig. 5h**). Thus overall, we found weak
345 signals for figure-ground and border-ownership modulation, which moreover depended strongly on specific
346 texture condition, across the three mouse visual areas surveyed.

347
348 Mean time courses of responses across the population to figure, ground, and border confirmed the strong
349 texture dependence of FGM and BOM signals (**Fig. 5i-k**). While there was clear enhancement in response
350 to the figure/border versus ground for the Cross condition starting at the earliest time point of response
351 (**Fig. 5i**), differences were much smaller for Iso and Nat conditions (**Fig. 5j, k**). Comparison of time courses
352 across areas revealed a more distinct response between figure and ground conditions in LM compared to
353 V1, and V1 compared to RL, with strong texture dependence of segmentation signals in all three areas
354 (**Extended Data Fig. 3**).

355 356 **Neural decoding of figure position mirrors behavioral performance**

357
358 The neural data so far shows a clear lack of texture-invariant segmentation signals in mouse visual areas
359 V1, LM, and RL (**Figs. 4, 5**). This is consistent with the mouse's inability to generalize segmentation across

360 textures for the Nat condition (**Fig. 2e, f, i1**). However, the mouse was behaviorally able to generalize
361 segmentation in the Cross and (to a lesser extent) Iso conditions (**Fig. 2i1**). To what extent can the neural
362 signals in mouse visual cortex explain this pattern of behavior?

363
364 To address this, we quantified how well we could read out the position of a figure on a given trial using a
365 linear decoder of neural responses. Within a single condition (Cross, Iso, Nat), decoding position would be
366 trivial for a single stimulus pattern: a set of ON cells with localized receptive fields like the hypothetical unit
367 in **Fig. 3d** (right) would be able to solve this task, as long as cells respond reliably and differentially to stimuli
368 with different figure locations. How well could a decoder trained on the mouse's neural responses
369 *generalize* segmentation across textures within a class? For each of the three conditions, we pooled trials
370 for the two orientations/patterns, and then trained a least squares linear regression model using 50/50 cross
371 validation over trials (**Fig. 6a**). **Fig. 6b** shows decoded versus actual figure position for varying numbers of
372 cells; for convenience, we decoded azimuth position. Decoding improved monotonically with the number of
373 cells used.

374
375 We quantified decoding performance as the variance in the azimuth position explained by the linear model
376 (**Fig. 6c**). Using electrophysiology data, we found that on average neural decoding was best for the Cross
377 condition ($r^2=0.89$ for 200 cells), followed by Iso ($r^2=0.53$ for 200 cells), and then Nat ($r^2=0.09$ for 200 cells).
378 This dependence of position decoding on texture condition (Cross > Iso > Nat) matched the ranking
379 observed in the behavioral performance of animals on the generalization stage of the figure localization
380 task (**Fig. 2i1**). In particular, variance explained was close to zero for the Nat condition.

381
382 Using imaging data, we found the same qualitative pattern, though overall decoding performance was
383 worse than that obtained from electrophysiology data for the same number of neurons (**Fig. 6d**), likely due
384 to the fact that the calcium signal is significantly noisier [34].

385
386 We next examined decoding performance for each of the conditions as a function of visual area. For both
387 the Cross and Iso conditions, decoding was best for LM followed by V1 and RL (for N = 120 cells: Cross:
388 LM > RL: $p < 10^{-4}$, LM > V1: $p < 10^{-4}$, V1 > RL: n.s.; Iso: LM > RL: $p < 10^{-4}$, LM > V1: $p < 10^{-4}$, V1 > RL: $p < 10^{-4}$,
389 rank sum test) (**Fig. 6e**). The same relationship was observed with imaging data (**Fig. 6f**).

390 391 **Mouse segmentation is well-modeled by a deep network**

392
393 How could neural signals in the mouse support linear decoding of figure position in the Cross and Nat
394 conditions despite lack of explicit figure-ground and border-ownership cells? To address this, we tested
395 different neural encoding models for how well they could explain the observed decoding performance. We
396 first simulated a population of 25,000 simple cells with varied receptive field size, location, orientation,
397 preferred phase, and spatial frequency (see Methods). Each unit consisted of a linear filter followed by
398 linear rectification and additive Gaussian noise (**Fig. 7a**). We refer to this model as the “feedforward LN
399 model.” We attempted to decode figure position from this model using the same procedures as we used for
400 analyzing the neural data (**Fig. 6a**). Surprisingly, we found that we could robustly decode figure position for
401 the Cross condition, though not for the Iso and Nat conditions (**Fig. 7b**). It is widely assumed that figure-
402 ground segregation (i.e., detecting the location of the square in the displays in **Extended Data Movie 2**)
403 cannot be accomplished through purely local linear filters. How could a simple feedforward LN model
404 decode figure position when the local stimulus at the figure center is identical to that of ground after pooling
405 across conditions? We realized that to decode figure position, one need not rely on signals from the center
406 of the figure; instead, one can use signals at the edges, and simple cells can readily localize orientation
407 discontinuities such as those present in the Cross condition. This underscores an important point: the Cross
408 stimulus completely fails as a behavioral marker for a nonlinear figure-ground segmentation process (see
409 also [35]).

410
411 We next modeled orientation-dependent surround modulation, a previously reported non-linear interaction
412 in mouse visual cortex [15, 16, 33]. To simulate orientation-dependent surround modulation, we added a
413 divisive term such that $response = response_{feedforward} / (1 + \beta * pearson(\vec{V}_{in}, \vec{V}_{out}))$, where
414 $pearson(\vec{V}_{in}, \vec{V}_{out})$ is the correlation between the mean orientation energy within a cell's receptive field (\vec{V}_{in}),

415 compared to that in the surround (\vec{V}_{out}). This surround model behaved similarly to our feedforward LN
416 model, failing to capture the texture dependence of the neural and behavioral data from the mouse (**Fig.**
417 **7c**). This is not surprising, since the nonlinear interaction in this model depends on an orientation
418 discontinuity, which was absent from the Iso condition. These results held generally across a range of noise
419 levels (**Extended Data Fig. 5**).

420
421 Finally, we hypothesized that while orientation-dependent surrounds might be an insufficient nonlinearity to
422 explain the mouse's behavioral and neural data, a deep convolutional network (DCN) trained on object
423 recognition might develop many nonlinearities useful for performing the figure localization task. For
424 example, common objects in cluttered scenes can resemble either the Cross or Iso conditions. We ran our
425 stimuli through Vgg-16 pre-trained on ImageNet to perform object classification (**Fig. 7d**) and analyzed
426 responses in convolutional layers 1-5 (**Fig. 7e**) [36]. We then tested decoding performance exactly as for
427 the feedforward LN and surround models by randomly drawing subsamples of cells from a given layer. The
428 performance of the DCN matched the mouse's neural and behavioral data well: performance was best for
429 Cross, followed by Iso, and then Nat, with this preference emerging in mid to late layers of the network.
430 These results held generally across a variety of noise levels (**Extended Data Fig. 6**). Thus a DCN replicated
431 the rank ordering of the mouse's behavior and neural decoding performance (Cross > Iso > Nat). This
432 suggests the mouse visual system may use similar nonlinear interactions as in a feedforward deep network
433 to accomplish object detection.

434 435 **Discussion**

436 We have shown that mice and primates segment visual input using fundamentally different strategies.
437 Unlike primates, mice are unable to detect figures defined purely by opponent motion, and hence mouse
438 segmentation behavior is strongly dependent on texture cues (**Fig. 7f**). Indeed when mice were forced to
439 detect figures defined purely by opponent motion for a limited number of patterns, they adopted a strategy
440 of brute force pattern memorization (**Fig. 2e-g**). The strong texture dependence of mouse object detection
441 behavior was consistent with neural signals recorded in mouse visual areas V1, RL, and LM, and could be
442 explained by a simple feedforward deep network model lacking explicit segmentation capabilities (**Fig. 7g**).

443
444 When we tested three additional species, the macaque, mouse lemur, and treeshrew, using the same
445 paradigm, we found that only the two primate species could perform segmentation using opponent motion.
446 It was especially surprising that the mouse lemur, a tiny (~60-80 g) prosimian primate species [37], could
447 segment purely motion-defined figures well above chance, while the treeshrew (~120-200 g), an animal
448 with a much more highly developed visual system than the mouse [38-40], could not. Overall, our findings
449 reveal a fundamental difference in the computational strategy used by mice versus primates for visual
450 segmentation and suggest that surface perception from accretion-deletion may be a capability unique to
451 primates. We believe this is highly significant because visual surface representation is a fundamental step
452 in primate visual processing [1, 21, 41], and accretion-deletion (**Fig. 1b, c**) has been recognized since the
453 seminal work of J.J. Gibson as *the most powerful cue* supporting visual surface representation [23]. In
454 particular, among all cues to surface organization, accretion-deletion is unique in its (i) high reliability,
455 occurring only at true object borders and never at internal texture borders, (ii) robustness to object texture,
456 and (iii) computational power, supporting not only segmentation but also invariant tracking without requiring
457 any learning or prior experience [21].

458
459 We were inspired by previous rodent behavioral studies that have sought to carefully characterize the visual
460 capabilities of mice and rats, testing behaviors such as transformation-tolerant object recognition and
461 natural scene discrimination [42-47]. In particular, consistent with our finding that mice cannot detect
462 naturalistic figures defined by opponent motion, Keyser et al. found that rats could not learn to detect a bar
463 defined by a grid of Gabor patches moving counterphase to ones in the background [42] (note, however,
464 their stimulus did not contain accretion-deletion, the cue we were especially interested in for reasons
465 explained in the previous paragraph). Overall, our results add to a growing body of work showing that rodent
466 and primate vision differ in essential ways beyond visual resolution.

467
468 Our finding that figure-ground signals exist in mouse visual cortex (**Fig. 3e, f, Fig. 4, Fig. 5d, i**), but are
469 strongly dependent on texture (**Fig. 4, Fig. 5a, b, e, i-k**), is consistent with previous studies [29-31, 48].
470 Optogenetic perturbation studies have further demonstrated that these signals are behaviorally relevant for

471 figure detection [29] and require feedback [29, 31]. Thus overall, it seems clear that mouse visual cortex
472 shows orientation-dependent surround modulation [15, 16] which can support texture-based figure-ground
473 segmentation behavior. However, importantly, our results show that mice lack a general, texture-invariant
474 mechanism for surface segmentation, unlike primates.

475
476 The inability of mice to detect figures defined purely by opponent motion was rather surprising, as cells
477 selective for local motion (distinct from global retinal image drift due to fixational eye movements) have
478 been reported in the retinas of both rabbits and mice [49, 50]. How could a signal present in the retina not
479 be used by the animal for segmentation? First, in our experiments, figure and ground moved exactly in
480 counterphase, a condition that the retinal object-motion cells are unable to detect [50]. Furthermore, retinal
481 studies have generally used sinusoidal gratings, and it remains unclear how responses of retinal object-
482 motion cells might generalize to arbitrary textures such as those used in our Nat task. Finally, we note that
483 object localization in the Nat task required perception of *differential* motion cues. We confirmed that mice
484 were readily able to detect the same moving figures against stationary backgrounds (**Extended Data Fig.**
485 **1a**). It is possible that retinal object-motion cells are adapted for this latter condition—whose handling may
486 be sufficient to ensure mouse survival—while different evolutionary pressures led to emergence of a more
487 sophisticated segmentation mechanism in primates.

488
489 The distinction between mouse and primate segmentation behavior has an intriguing parallel in the
490 difference between deep network and human object classification behavior. In recent years, deep network
491 models of vision have started to achieve state of the art performance on object classification tasks.
492 However, the lack of an explicit segmentation process in these networks leads to susceptibility to
493 adversarial examples in which noise patterns that lack any surface structure are classified as objects with
494 high confidence [51, 52]. Furthermore, recent work has shown that deep networks, unlike humans, fail to
495 exploit global image features such as object boundaries and shape to perform classification, instead relying
496 much more strongly on texture features [53, 54]. Thus it is clear that a major difference between deep
497 networks and the primate visual system lies in their segmentation capabilities. Our finding of strong texture-
498 dependence in mouse segmentation behavior suggests that mice adopt a visual strategy more similar to
499 deep networks than primates do (**Fig. 7g**; see also [55]).

500
501 Objects are the fundamental building blocks of a primate's model of the world. An object is first and foremost
502 a *moveable* chunk of matter, i.e., a segmentable surface. The remarkable trajectory of the human species
503 has been variously attributed to language, tool use, upright posture, a lowered larynx, opposable thumbs,
504 and other traits [56]. We close with a speculation: it may not be entirely implausible that possession of
505 machinery for cue-invariant visual object segmentation played a role in setting the human trajectory. By
506 enabling creation of a rich and accurate physical model of the world inside the primate brain, this perceptual
507 machinery, seemingly less developed in all non-primate species tested so far including both mice and
508 treeshrews, may have supplied the foundation for subsequent human capabilities requiring a hyper-
509 accurate model of objects in the world—including tool use, causal understanding, and general intelligence.

510
511

512 **Methods**

513

514 **Animal statement**

515 The following animals were used in this study: adult mice 2–12 months old, both male and female; adult
516 treeshrews 7–18 months old, both male and female; adult mouse lemurs 2–3.5 yrs, both male and female;
517 and adult macaques 3 and 7 yrs old, male. All procedures on mice, macaques, and tree shrews were
518 conducted in accordance with the ethical guidelines of the National Institutes of Health and were approved
519 by the Institutional Animal Care and Use Committee at the California Institute of Technology.

520

521 Mouse lemur experiments were in accordance with European animal welfare regulations and were reviewed
522 by the local ethics committee (“Comite d’éthique en expérimentation animale No. 68”) in Brunoy, France,
523 by the ethics committee of the University of Geneva, Switzerland and authorized by the French “Ministère
524 de l’éducation nationale de l’enseignement supérieur et de la recherche.”

525

526 **Transgenic animals**

527 For imaging experiments, we used a cross between a CamKII::tTA mouse (JAX: 00310) with a
528 tetO:GCaMP6s (JAX: 024742) to target expression to cortical excitatory neurons. For electrophysiology
529 experiments we used a Thy1::GCaMP6s 4.12 (JAX: 025776). Behavioral experiments were carried out with
530 a combination of Thy1 and C57Bl6 animals. We back crossed all lines to C57Bl6.

531

532 **Surgical procedures**

533 The cranial window and headplate procedures were based on Wekselblatt et al. [57] with some
534 modifications as described below.

535

536 **Headplate surgery.** For both electrophysiology and imaging experiments, a stainless steel headplate was
537 attached to the animal’s skull in a short procedure. Animals were anesthetized using isoflurane (3%
538 induction; 1.5%–2% maintenance) in 100% O₂(0.8–1.0 l/min) and positioned in a stereotax using earbars
539 placed just below the ear canal for stability. The animals were given subcutaneous injections of the
540 analgesic Ketoprofen (5 mg/kg) and 0.2 ml saline to prevent postoperative dehydration. Body temperature
541 was maintained at 37.5°C by a feedback-controlled heating pad; temperature and breathing were monitored
542 throughout surgery. Sterilized instruments and aseptic technique were used throughout. Sterile ocular
543 lubricant (Puralube) was applied at the beginning of each surgical procedure. Scalp hair was removed using
544 an electric shaver, and the surgical site was cleaned using a combination of dermachlor and chlorohexidine.
545 A scalp incision was made using #3 scissors (FST) and the periosteum was pulled back using forceps. The
546 back neck muscles were retracted to make room for a either an 8 or 10 mm circular opening headplate
547 which was affixed to the skull using either metabond (Parker) or dental acrylic (OrthoJet). A combination of
548 vet bond and cyanoacrylate-based glue was applied to the skull to both protect the skull surface from
549 infection and also to provide a clear surface through which to perform widefield imaging to identify cortical
550 visual areas in electrophysiology experiments.

551

552 **Craniotomy surgery/window implantation.** After allowing the animal to recover at least 1 week from the
553 headplate surgery a craniotomy procedure was performed to either allow for acute implantation of an
554 electrode or to install a glass coverslip for chronic imaging.

555

556 Animals were anesthetized using isoflurane (3% induction; 1.5%–2% maintenance) in 100% O₂ (0.8–1.0
557 l/min) and positioned in the stereotaxic frame affixed by the headplate attached previously. The animals
558 were given subcutaneous injections of the analgesic Ketoprofen (5 mg/kg) and 0.2 ml saline to prevent
559 postoperative dehydration. Body temperature was maintained at 37.5°C by a feedback-controlled heating
560 pad; temperature and breathing were monitored throughout surgery. Sterilized instruments and aseptic
561 technique were used throughout. Sterile ocular lubricant (Puralube) was applied at the beginning of each
562 surgical procedure.

563

564 For imaging experiments a 4–5 mm craniotomy was cut out centered at +0.5 mm from lambda and +2.75
565 mm from midline on the right hemisphere. Care was taken to minimize bleeding, and any bleeds in the skull
566 during drilling were covered with wet gelfoam (Pfizer) until they resolved. After careful removal of the bone
567 flap, a durotomy was performed and the exposed brain was covered in a 1:1 mix of artificial dura (Dow

568 Corning 3-4680). A sterile 4-5 mm coverslip was then pressed into the opening and sealed in place using
569 a combination of cyanoacrylate-based glues. The remaining parts of exposed skull in the headplate well
570 were then covered with black dental acrylic for light blocking purposes and to prevent infection.

571
572 For electrophysiology experiments, retinotopic mapping was performed prior to performing any craniotomy,
573 resulting a vasculature and field sign map to identify vasculature landmarks corresponding to either V1, LM,
574 or RL. Once such landmarks had been identified a small (<1 mm) craniotomy was performed on the morning
575 of each experiment. The craniotomy, was sealed with KwikSil (WPI) and animals were allowed to recover
576 for at least 3 hours to subsequent recording experiments.

577 578 **Visual stimuli**

579 Visual stimuli were presented on a 32 inch monitor (Samsung 32 inch lcd screen; 42 × 74 cm); linearized
580 by eye to correct for gamma (mean luminance 50 cd/m²), oriented tangentially 18 cm from the mouse's left
581 eye (figure size: ~30°) for all experiments except those shown in **Fig. 3f** for quantifying reproducibility which
582 were carried out at 27 cm from the eye (**Fig. 3b**).

583
584 For all experiments (electrophysiology and imaging), attempts were made to re-center the monitor on the
585 general location of the receptive fields across the population to maximize the number of distinct figure and
586 ground trials in a given experiment.

587
588 **Figure ground/border ownership stimulus.** The stimuli used to characterize the figure-ground and
589 border-ownership modulation of each stimulus consisted of a sinusoidal grating of 0.06 cpd, oriented at
590 either 45 or 135°. The figure was 27° in size, and all the horizontal positions surveyed spanned 45° (~3°
591 shift per position). The total elevation positions surveyed varied across 23° (~3° shift per position). For the
592 Iso and Nat conditions where the texture defining the figure was identical within a given condition, the figure
593 was generated at each position using the same texture that was there in the background as a new
594 foreground image, thus creating a “pop out” effect for the stimulus (see **Extended Data Movie 2**).

595
596 **RF mapping.** Receptive fields were mapped for neurons under 2-photon using an isolated drifting Gabor
597 patch stimulus: a patch of ~6° containing a drifting Gabor appeared in one of three different orientations
598 (45°, 180°, 315°) and 2 directions at a random position. We repeated this procedure for 9 repeats and at
599 16 by 9 positions and then collapsed responses across all trial types to compute a spatial PSTH. We then
600 fit a 2D Gaussian to the response and classified neurons as having a significant receptive field fit if the
601 goodness of fit exceeded the goodness of fit for at least 99 out of 100 bootstrapped trials where spatial
602 location was shuffled across trials.

603
604 For electrophysiology, we used a sparse noise stimulus with light and dark spots to map receptive fields.
605 We computed separate PSTHs for ON and OFF subunits, fit 2D Gaussians, and determined significance
606 using the same procedure as above using shuffled bootstraps.

607
608 We limited our analysis for figure-ground and border-ownership modulation in **Fig. 5** to neurons that had
609 significant receptive field fits and used the central position of that fit as the neuron's inferred RF center. For
610 population neural decoding analyses, we used all neurons regardless of receptive field fits.

611 612 **Mouse behavior**

613 We trained sixteen adult mice aged 8-12 weeks (4 x C57bl/6; 12 x Thy1-Gcamp6s mice) co-housed 4 to a
614 cage with access to a running wheel. Mice were housed under reversed light-dark cycle (dark cycle from
615 10 am to 9 pm), and training was performed during the dark cycle. Mice were water restricted before, during,
616 and after training days. The weight of the animals was continuously monitored to make sure no mouse
617 dropped below 85% of their pre-water-restriction weight.

618
619 Mice were trained for 6 days/week, with a single session per day. Each mouse performed around 200 to
620 450 trials per day depending on the behavioral task. All training was carried out a commercially available
621 touchscreen operant chamber in a sound-attenuating box (Bussey-Saksida Touch Screen Operant
622 Chambers/Lafayette Instrument).

623 Visual stimuli were presented on a touchscreen monitor (24 x 18.5 cm, W x H) (Lafayette Instrument). In all
624 experiments, size of figures is given in pixels as the animals were freely moving and thus no estimate can
625 be made of the size of the figure in visual degrees during the behavior as it will vary depending on the
626 animal's position. We generated stimulus movies of size 100 pixels high x 200 pixels wide, with square
627 figures centered at either 50 or 150 pixels and square side of 50 pixels. Both figure and background moved
628 out of phase with a sine wave of amplitude 14 pixels.

629 One side of each chamber was a touch screen monitor, and on the other side was a reward tray with a
630 water spout. Mice had to touch the left or right side of the screen based on the location of the figure. The
631 water delivery was cued by a white light illuminating the spout. An IR sensor detected when the mouse
632 collected the water reward and determined the start of the next trial. For incorrect responses, a white noise
633 audio stimulus was played for three seconds and followed by a ten second timeout. After the time-out period,
634 the stimulus of the identical trial was presented again until the mouse made the correct touch. This shaping
635 procedure can prevent biases in the behavior. All trials following the first incorrect response were not
636 included in the analysis, as other similar studies have done [30, 58].

637
638 We first trained the animal to become familiar with the touch screen and the task rules through a luminance
639 shaping procedure. During this procedure, the mice learned to discriminate a white square moving on a
640 black background and touch the side of the screen that displayed the square. In this task, they also learned
641 that a correct touch was associated with a water reward. The mice started their subsequent training steps
642 after reaching 80% on this task.

643
644 **Standard 2-pattern discrimination.** For these experiments, there were 4 possible stimuli (2
645 orientations/natural noise x 2 positions) (**Fig. 2b, Fig. 2i1, Extended Data Fig. 1c1**). Gratings were
646 presented at two orientations, 45° and 135° (0.25 cycles/cm).

647 **5-pattern discrimination.** Animals were tested on this after training on the 2-pattern task (**Fig. 2i1**). The
648 stimuli were the same as for the 2-pattern discrimination, except we introduced 5 novel orientations or
649 naturalistic noise patterns. For gratings, the 5 novel orientations were 22.5°, 67.5°, 90°, 112.5°, 157.5°. For
650 the Nat condition, we used 5 new natural noise patterns that shared the same Fourier statistics.

651
652 **30 natural textures.** After finishing training on 2- and 5-pattern discrimination tasks, mice were trained on
653 30 natural textures in Iso and Cross configurations (**Extended Data Fig. 2**). We first tested them on 7
654 textures (set A) to measure the baseline performance. We then trained them on 30 Iso and Cross natural
655 textures (set B). After training, they were tested on 7 textures again (set A). For experiments involving
656 natural textures we randomly selected textures from the describable textures dataset (DTD;
657 <https://www.robots.ox.ac.uk/~vgg/data/dtd/>). We converted all textures to grayscale.

658 **Grating to noise shaping.** We generated a gradual morph stimulus that changed oriented gratings into
659 naturalistic noise, with the goal being to train the animals to use the motion cue (**Fig. 2e**). In total, there
660 were 10 stages in this shaping procedure. At each stage, the stimulus represented a weighted sum of the
661 grating and noise, and the weight was changed 10% from the previous stage for both grating (decrease)
662 and noise (increase). For example, in stage 3, the weight assigned to grating was 70% and the weight for
663 noise is 30%. By stage 10, the weight for grating was 0% and for noise was 100%. At each stage, figure
664 texture was flipped vertically and presented at a random height (top, middle, down) to discourage use of
665 local cues in the task.

666
667 **Static version of tasks.** We tested mice on static versions of all conditions where only a single frame of
668 the stimulus was presented (**Fig. 2g, h**). The static stimulus was the frame that had the largest phase
669 difference between figure and background.

670
671 **New texture task.** After mice learned the grating to noise task, they were tested on 7 unseen new natural
672 textures (**Fig. 2f**).

673
674 **Background static natural textures.** We trained the mice to detect a figure moving horizontally on a static
675 background (**Extended Data Fig. 1a**).

676

677 **Treeshrew behavior**

678 Four adult treeshrews (3 male, 1 female; age: 7-18 months old) bred and raised at Caltech were trained to
679 perform the figure-ground segregation task. Animals were singly housed in a 12 hr:12 hr light:dark cycle.
680 They were not food or water restricted, but free access to water was limited during the 4 hours prior to
681 training each day. Training was performed during the light cycle in a custom-made behavioral arena (30 x
682 30 x 25 cm) containing three optical lickports (Sanworks) situated in a custom-built behavior box. Drops of
683 100% apple juice rewards were provided upon poking at the appropriate lickports and in some cases for
684 trial initiation. Images were presented on a Sunfounder 10.1" Raspberry Pi 4 (1280 x 800) screen and
685 controlled using Bpod hardware and Python software. After an initial shaping step in which animals learned
686 to use the lickports in a luminance detection task (2-3 days), training for Cross, Iso, and Nat conditions was
687 performed for 4 consecutive days (5 orientations or naturalistic textures), with generalization test sessions
688 on the fifth day (2 different orientations or naturalistic textures) (**Extended Data Fig. 1b, c2**). We reversed
689 the number of train/test patterns compared to what was used for the mice (**Fig. 2i1**) because we reasoned
690 that animals might be more likely to generalize if given more patterns for training.

691 **Mouse lemur behavior**

692 Four adult mouse lemurs (3 male, 1 female; age: 2-3.5 yrs) bred and raised in the "Mouse Lemur Platform"
693 (authorization number E-91-114-1) of the "Museum National d'Histoire Naturelle" in Brunoy, France (UMR
694 MECADEV CNRS/MNHN 7179) were trained to perform the figure-ground segregation task. Animals were
695 co-housed 2-3 per cage in a reversed long-day (14:10 light:dark) cycle. They were food restricted, with their
696 body weight maintained above 60g, but had free access to water. Training was performed during the dark
697 cycle in a custom-made behavioral arena (20 x 20 x 30 cm) containing three optical lickports (Sanworks)
698 situated in a sound-attenuating box. Drops of liquid food rewards (standard food mixture composed of
699 banana, cereal, milk and egg) was provided upon poking at the appropriate lickports. Images were
700 presented on a Dell P2414H (1920 x 1080, 60Hz) screen and controlled using Psychopy and Matlab
701 software. Training and testing followed the same paradigm as for treeshrews (**Extended Data Fig. 1b, c3**).

702 **Macaque behavior**

703 Two head-fixed rhesus macaque monkeys were trained to indicate whether a square was on the left or right
704 side of a screen. Movie stimuli were shown on an LCD screen in pseudo-random succession for 2 seconds
705 ON time each, without any OFF period. The stimuli were shown across the full screen (23° in height and
706 37° width) and contained a square of 9° length on either the left or right side. Monkeys received a juice
707 reward for fixating within the 9° square region for at least 1 second. Eye position was monitored using an
708 infrared eye tracking system (ISCAN). Each monkey performed only one or two sessions of one or two
709 hours each. In the beginning of the first session, behavior was shaped by training the monkeys on the
710 luminance squares only until they reached 90% correct performance. Prior to this, the monkeys had been
711 trained only to fixate. Both monkeys learned this within the first session and subsequently performed the
712 task with all other stimuli presented in pseudo-random succession (**Extended Data Fig. 1c4**). Stimuli
713 included static and moving luminance squares, cross-orientation and iso-orientation gratings and natural
714 textures.

715 For offline analysis, we computed the percentage of correct trials where monkeys were fixating the correct
716 square location for at least 1 second, including only trials where monkeys were looking at the screen and
717 not closing the eyes.

718 **Wide-field imaging**

719 Prior to all electrophysiological and imaging experiments, a reference vasculature image and field-sign map
720 was acquired under a custom-built widefield epi-fluorescence microscope. The microscope consisted of
721 two Nikon 50 mm f1.4 lens placed front to front with a dichroic, excitation, and emission filters (Semrock) in
722 between. Light was delivered via a blue LED light source (Luxeon Star) and images were acquired with a
723 CMOS camera (Basler). Images were acquired at 10 Hz and were triggered on every 3rd frame of a 30 Hz
724 retinotopic mapping stimulus (drifting bar; trial period of 0.1 Hz) to ensure proper timing between stimulus
725 and acquisition. Retinotopic mapping stimulus consisted of a drifting 10° bar of binarized 1/f noise [57],
726 which cycled with a period of 0.1 Hz. Elevation and azimuth maps were computed using a Fourier
727 decomposition of the stimulus and plotting preferred phase at the stimulus frequency [59].

728

733 **Two-photon imaging**

734 We began imaging sessions ~2 weeks after surgery. We used a resonant, two-photon microscope
735 (NeuroLabware, Los Angeles, CA) controlled by Scanbox acquisition software (Scanbox, Los Angeles, CA).
736 Imaging was through a 16x water immersion lens (Nikon, 0.8 NA) at an acquisition rate of 15.6 Hz at depths
737 ranging from 150 to 250 μm from the surface corresponding to layer 2/3. Mice were allowed to run freely
738 on a spherical treadmill (styrofoam ball floated with air).

739
740 We ran a minimum of 7 stimulus conditions in all sessions (RF mapping + 6 conditions) with a short (<3
741 min) break between each imaging session. Most sessions lasted less than 75 minutes. For analysis, all of
742 the movies from each session were aligned to a common mean image using a non-rigid registration pipeline
743 (Suite2P). Suite2P was further used to extract cell locations and perform subsequent deconvolution to
744 recover spike estimates from calcium signals. Results of cell extraction were inspected manually using the
745 Suite2P GUI and any outliers (e.g., clearly non-neuronal or artifactual shapes) were discarded.

746
747 **Matching cells across days.** Cells were tracked across days by first re-targeting to the same plane by eye
748 such that the mean fluorescence image on a given day was matched to that on the previous day, with online
749 visual feedback provided by a custom software plugin for Scanbox. Then, a registration correction was
750 computed from one day to the other such that the two planes were aligned. After extracting cell identities
751 and cell filter maps for both days independently, a Jacard index $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$ was computed for a given
752 cell filter on day 1 with all extracted filters on day 2. A Jacard index > 0.5 was considered the same cell on
753 the next day.

754 **Electrophysiology**

755
756 Electrophysiology experiments were carried out with an acutely-inserted 64 channel silicon probe [60]
757 attached to a 4-axis manipulator (Siskiyou), that was amplified through a 64 channel headstage (Intan).
758 Signals were digitized by an open-ephys acquisition box (Open-ephys) and aligned to stimulus frames
759 through the use of a photodiode. The probe was lowered into the brain at 5 $\mu\text{m}/\text{sec}$ and allowed to settle
760 for a minimum of 5 minutes before experimental stimuli were presented. Animals were head-fixed for a
761 maximum of 2 hours in any given experiment. Stimuli were presented as outlined in the Visual Stimuli
762 section of the Methods.

763 **Analysis**

764 All analyses were performed using custom scripts written in MATLAB (Mathworks) or Python using NumPy,
765 SciPy, Pandas, seaborn, sklearn, and Matplotlib [61-65].

766
767
768 **Trial-based response.** For all trial-based analysis, we quantified the response for a given trial as the mean
769 spike count 50-250 ms post trial onset. For imaging experiments, we use the deconvolved calcium trace
770 where the response value was set to the mean across all frames of the trial.

771
772 **Positional decoding using linear regression.** To quantify the amount of information present about figure
773 position, we decoded the azimuth bin (positions 1-16) of each trial from a population of neurons using a
774 ridge regression model and 50/50 cross validation; results averaged across 100 iterations are reported for
775 all data and modeling. To quantify the extent to which a single linear model could account for position
776 across both orientations or textures, we pooled trial types within a stimulus condition (both
777 orientations/textures for Cross/Iso/Nat).

778
779 Beta values (penalty term) for the ridge regression were computed with a leave-one-out cross-validation
780 approach using the RidgeCV function from sklearn. As the beta values will be dependent on the number of
781 regressors (neurons) in the model, this entire procedure was repeated for varying numbers of neurons to
782 compute the decoding performance as a function of number of neurons in the decoder. All values reported
783 for decoding correspond to the mean variance explained by the model over 100 iterations of the above
784 procedure, and error bars correspond to 95% confidence intervals.

785
786 **Computing FGM and BOM.** We limited analysis of figure-ground modulation (FGM) and border-ownership
787 modulation (BOM) to electrophysiologically-recorded neurons that satisfied two criteria: i) they showed a

788 statistically significant receptive field fit, and ii) the receptive field center was limited to a central portion of
 789 the screen (central 15° azimuth and 10° elevation). This second receptive field position criterion was to
 790 ensure a reasonable number of figure and ground or left and right trials with which to compute the FGM or
 791 BOM indices.
 792

793 For each of the three conditions, we defined a figure ground modulation index as $FGM = \frac{(R_{Fig} - R_{Back})}{(R_{Fig} + R_{Back})}$, where
 794 R_{Fig} is the mean response across the two patterns for the condition within the figure zone, defined as the 2
 795 x2 (10° x 10°) grid of locations centered on the cell's receptive field (**Fig. 5c**) and R_{Back} is the mean
 796 response across the two patterns for the condition in the background zone, defined as all grid locations with
 797 distance greater than 1.5 * the receptive field width from the receptive field center. We computed
 798 bootstrapped p-values using 500 shuffles where trial identity was randomized to establish a null distribution.
 799

800 We quantified border ownership selectivity in a similar way. We defined a border ownership modulation
 801 index as $BOM = \frac{(R_{Left\ border} - R_{Right\ border})}{(R_{Left\ border} + R_{Right\ border})}$, where $R_{Left\ border}$ is the mean response across the two patterns
 802 for the condition within the left border zone, and $R_{Right\ border}$ is the mean response across the two patterns
 803 for the condition within the right border zone (**Fig. 5f**). We computed the significance of the modulation
 804 using a bootstrapped distribution as above.
 805

806

807 Modeling

808

809 **Feedforward model.** We modeled neuronal responses using an LN model of simple cells, with the linear
 810 filter modeled by a Gabor function and linear rectification. We simulated responses to 100 different cell
 811 types using a classic simple cell model as a linear combination of receptive field with stimulus passed
 812 through a nonlinearity.
 813

813

814

$$814 \quad response_{feedforward} = f(g(x, y, \theta, \lambda, \gamma, \sigma, \omega) * stimulus)$$

815

816 where $g(x, y, \theta, \lambda, \gamma, \sigma, \omega)$ is a Gabor function:
 817

817

$$818 \quad g(x, y, \theta, \lambda, \gamma, \sigma, \omega) = e^{-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}} \cos\left(2\pi \frac{x'}{\lambda} + \omega\right), \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

819

820 with parameters sampled as follows: $\theta = \text{rand}(0, \pi)$, $\sigma = \text{rand}(2^\circ, 7^\circ)$, $1/\lambda = \text{rand}(0.05 \text{ cpd}, 0.3 \text{ cpd})$, $\omega =$
 821 $\text{rand}(0, \pi)$, $\gamma = 1$. $f(x)$ represents a linear rectification ($\max(0, x)$) to ensure positive rates and * represents
 822 the convolution operator.
 823

822

823 **Surround model.** We added a divisive term to the neural response computed from our feedforward LN
 824 model (**Extended Data Fig. 4**). This divisive term was not recurrent and instead can be most readily
 825 interpreted as a center-surround interaction that would arise in the feedforward inputs from thalamus to
 826 cortex.
 827

827

828

$$828 \quad response = \frac{response_{feedforward}}{1 + \beta * \rho(\vec{V}_{in}, \vec{V}_{out})}$$

829

830 where $\rho(\vec{V}_{in}, \vec{V}_{out})$ represents the Pearson correlation between the mean response of all neuron types
 831 within the receptive field ($< 2\sigma; \vec{V}_{in}$) and all neuron types outside the classical receptive field ($> 2\sigma \ \& \ < 5\sigma;$
 832 \vec{V}_{out}). The term $\rho(\vec{V}_{in}, \vec{V}_{out})$ represents the extent to which orientations inside the cell's receptive field match
 833 those outside of the receptive field for a given image. In the case of the Iso stimulus, $\rho(\vec{V}_{in}, \vec{V}_{out}) > 0$,
 834 yielding suppression, while in the case of Cross stimulus, $\rho(\vec{V}_{in}, \vec{V}_{out}) < 0$, yielding facilitation. We set the
 835 scaling factor β to 0.95 to ensure that the denominator never reaches zero.
 836

836

837 **Neural network model.** We analyzed decoding performance of intermediate layers of a deep network
838 (VGG-16) trained for classification on ImageNet [36]. We used the mean of the response magnitude to all
839 14 individual frames corresponding to a given stimulus to represent a given unit's response to a given trial
840 type. We extracted this response for each of the 5 max-pool layers in the network and computed decoding
841 curves using the same procedures as for the neural data.
842

843 **Noise factor.** To account for neural response variability, we added noise to all models proportional to the
844 mean of responses for the simulated network within each condition. To achieve this, we added zero-mean,
845 normally-distributed fluctuations to each trial with variance proportional to the mean response across
846 neurons followed by linear rectification to ensure positive outputs. Thus the final output of a population
847 would be equal to $response + N(0, noise_{factor} * pop_{mean})$. Sweeps across various noise factors are shown
848 in **Extended Data Figs. 5, 6**.
849

850 **Links to supplemental movies:**

851

852 **Extended Data Movie 1:** <https://youtu.be/LVRxtmrT168>

853 **Extended Data Movie 2:** <https://youtu.be/YyJ64ngzh8k>

854

855

856

857 **Acknowledgments:**

858 This work was supported by NIH (DP1-NS083063) and the Howard Hughes Medical Institute. We thank
859 Audo Flores and Daniel Wagenaar for technical support, Sotiris Masmanidis for supplying the silicon
860 recording probes, and David Fitzpatrick and Yong-Gang Yau for invaluable help setting up a tree shrew
861 colony. F.J.L. was supported by an Arnold O. Beckman postdoctoral fellowship and a Burroughs Wellcome
862 PDEP Award. D.H and C.L.A.H. were supported by the Swiss National Science Foundation
863 (310030E_190060) and the Human Frontiers Science Program (RGP0024/2016).

864

865 **Author contributions:**

866 F.J.L., L.L., and D.Y.T. designed the experiments. F.J.L. and L.L. collected mouse data, C.L.A.H. and D.H.
867 collected mouse lemur data, J.H. collected macaque data, and J.B.W, F.L., and F.J.L. collected treeshrew
868 data. F.J.L. and L.L analyzed all the data. F.J.L., L.L., and D.Y.T. interpreted the results and wrote the
869 paper, with feedback from other authors.

870

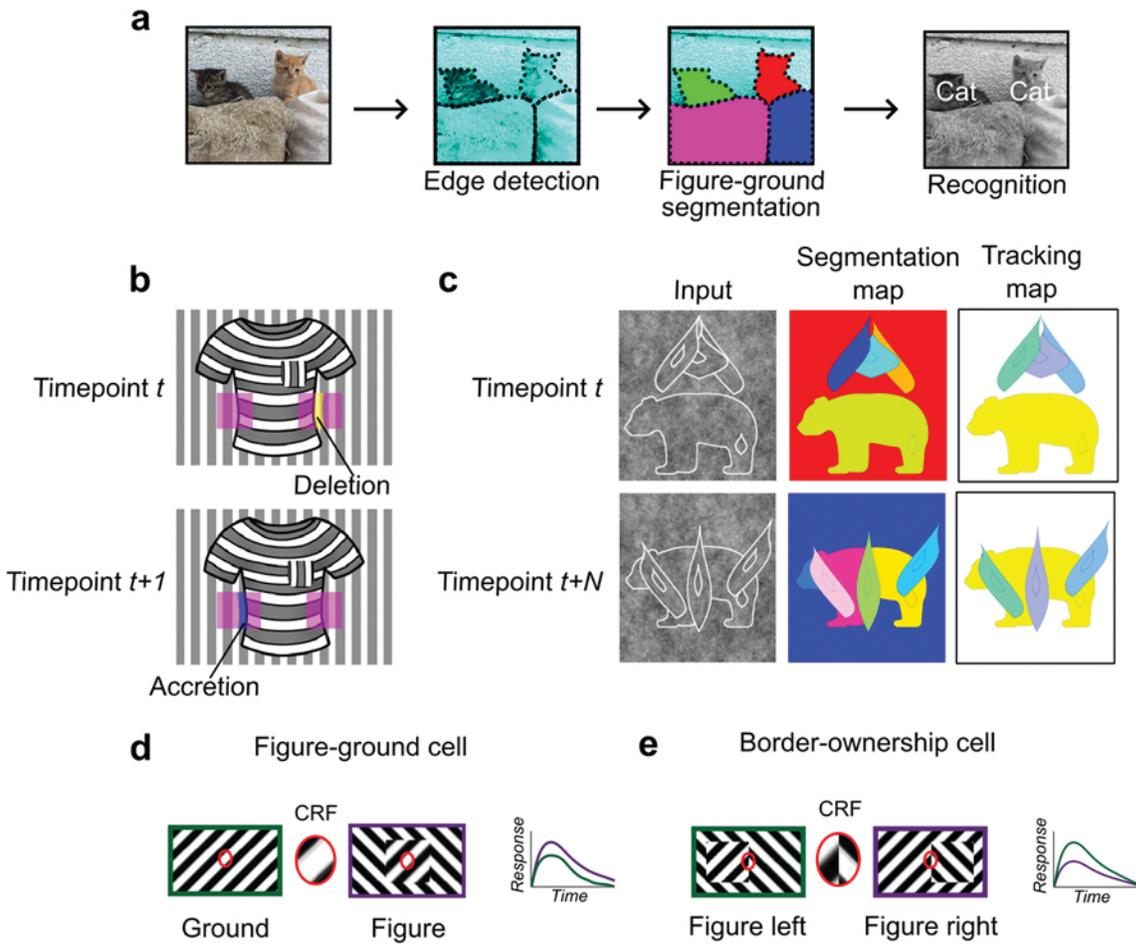


FIGURE 1

871 **Figure 1: Mechanisms for segmentation.**

872

873 a) Schematic representation of a hierarchy for visual perception. Figure-ground segmentation serves as a
874 key intermediate step preceding object recognition.

875

876 b) Accretion and deletion signals at borders induced by object motion provide a strong cue to distinguish
877 object vs. texture edges. As objects move differently from the background, accretion (left) and deletion
878 (right) of parts of the background will occur at object edges, providing local cues for object boundaries and
879 their owners. In contrast to accretion-deletion, texture (e.g., orientation contrast) is locally ambiguous: the
880 pocket does not constitute an object edge, even though it generates a sharp texture discontinuity.

881

882 c) Left column: two frames taken from a movie of three leaves descending over a bear. Accretion and
883 deletion signals support not only surface segmentation from a successive pair of image frames (middle
884 column), but also invariant object tracking across many successive image frames (right column), as
885 explained in detail in [21]. The key idea is that accretion-deletion enables surfaces to be *correctly linked*
886 between successive image frames, because the owner remains invariant between successive frames. In
887 (b), this corresponds to linking the surface of the shirt at time points t and $t+1$. When this process is carried
888 out across many successive time points, it enables correct linking of highly disparate object appearances,
889 such as the intact bear (top right) and the bear split into three pieces by occlusion (bottom left). Adapted
890 from [21].

891

892 d) Figure-ground modulation provides a neural mechanism for explicit segmentation. Here, a hypothetical
893 neuron's firing is selectively enhanced to a stimulus when it is part of a figure (purple) compared to ground
894 (green), even though the stimulus in the classical receptive field remains the same. A population of such
895 neurons would be able to localize image regions corresponding to objects.

896

897 e) Border-ownership modulation provides an additional neural mechanism for explicit segmentation. Here,
898 a hypothetical neuron's response is modulated by the relative position of a figure relative to an object edge.
899 In the example shown, the neuron prefers presence of a figure on the left (green) as opposed to figure on
900 the right (purple). A population of such neurons would be able to effectively trace the border of an object
901 and assign its owner.

902

903

904

905

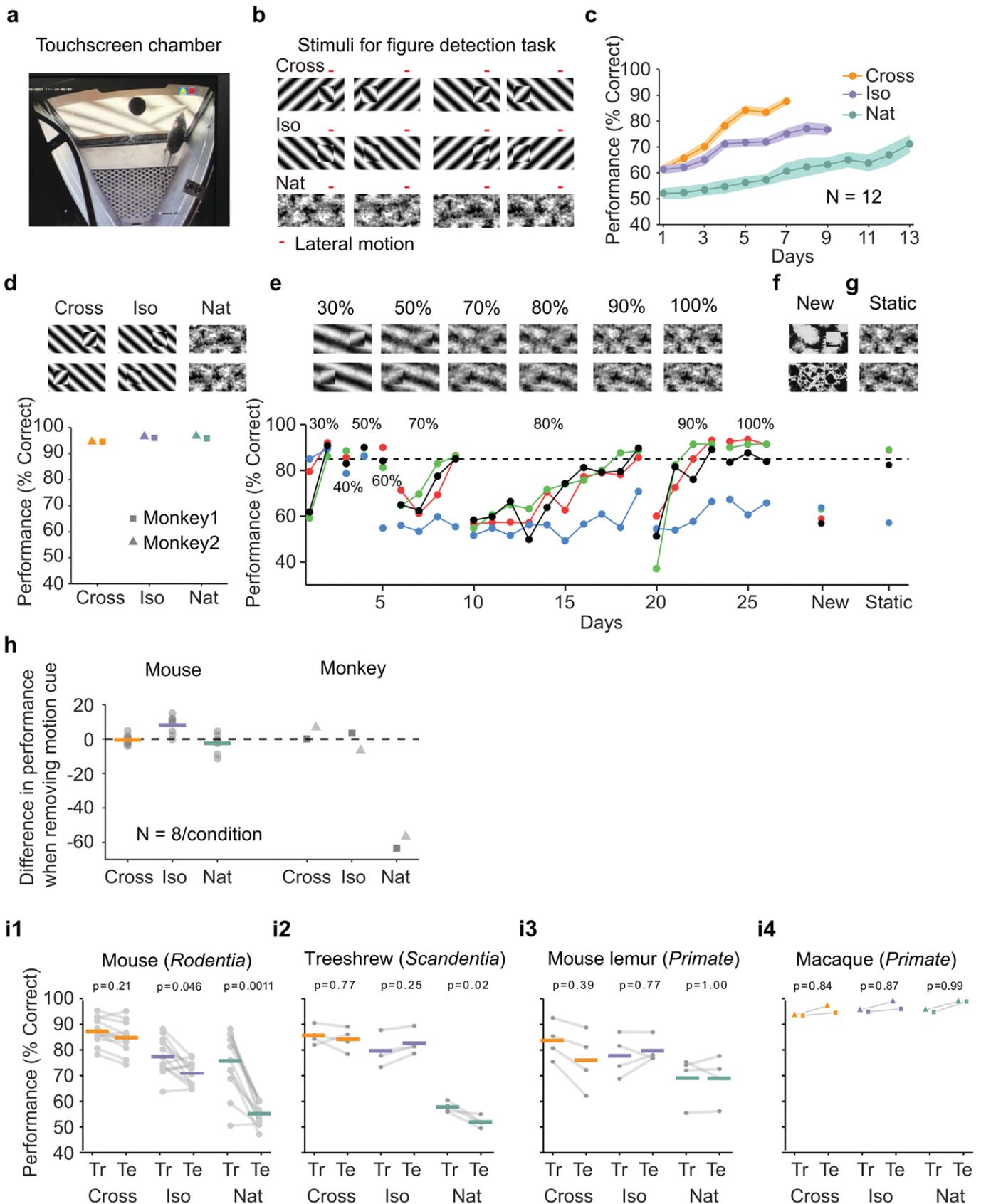


FIGURE 2

906 **Figure 2: Mouse segmentation behavior: mice use orientation contrast but not opponent motion to**
907 **distinguish figure from ground.**
908

909 a) Mice were trained in a touchscreen paradigm in which they were rewarded for touching the side of the
910 screen containing a texture- and motion-defined figure.
911

912 b) Mice were tested on three classes of stimuli: "Cross" where foreground and background patterns
913 consisted of orthogonal gratings, "Iso" where foreground and background patterns consisted of the same
914 orientation gratings, and "Nat" where foreground and background patterns consisted of naturalistic noise
915 patterns with 1/f spectral content. Initially, four training stimuli were used for each condition. Figure and
916 background oscillated back and forth, out of phase, providing a common opponent motion cue for
917 segmentation across all conditions; the movement range of the figure and background is denoted by the
918 red bar.
919

920 c) Mean performance curve for 12 mice in the Cross (orange), Iso (violet), and Nat (green) conditions,
921 where the task was to report the side of the screen containing a figure; in each session, one of a bank of
922 four possible stimuli were shown, as in (b). Shaded error bars represent sem.
923

924 d) Performance of two macaque monkeys on the same task. Monkey behavior, unlike that of mice, showed
925 no dependence on the carrier pattern, displaying high performance for all three conditions (Cross, Iso, Nat).
926

927 e) Teaching a mouse the Nat condition. Mice that could not learn the Nat version of the task could be
928 shaped to perform well on the task by a gradual training regimen over 20+ days. Using a gradual morph
929 stimulus (see Methods), animals could be slowly transitioned from a well-trained condition (Cross) to
930 eventually perform well on the full Nat task. Each circle represents one mouse.
931

932 f) Despite high performance on the four stimuli comprising the Nat task, performance dropped when mice
933 were exposed to new unseen textures, suggesting that they had not learned to use opponent motion to
934 perform the task.
935

936 g) Mice performed just as well on the Nat task even without the opponent motion cue, suggesting that they
937 had adopted a strategy of memorizing a lookup table of textures to actions, rather than performing true
938 segmentation in the Nat condition.
939

940 h) Left: Change in performance when the motion cue was removed on a random subset of trials. Mice
941 experienced no drop in performance in any of the conditions when static images were displayed instead of
942 dynamic stimuli, indicating they were not using motion information. Right: In contrast, monkeys showed no
943 performance drop in conditions where the figure was obvious in static frames (Cross and Iso), but showed
944 a marked drop in performance for the Nat condition where the figure is not easily resolved without the
945 motion cue.
946

947 i1) To confirm whether mice used an opponent motion cue in the various conditions, mice were trained on
948 an initial set of 4 stimuli (Tr; 2 sides x 2 patterns/orientations, as in (b)). After performance had plateaued,
949 they were switched to 10 novel test conditions (Te; 2 sides x 5 patterns/orientations). Animals mostly
950 generalized for Cross and Iso conditions but failed to generalize for the Nat condition ($p=0.0011$, ranksum
951 test), suggesting they were unable to use the opponent motion in the stimulus.
952

953 i2) Same as (i1) for treeshrews. Like mice, treeshrews failed to generalize for the Nat condition ($p=0.02$,
954 ranksum test).
955

956 i3, 4) In contrast, two primate species: mouse lemurs (i3) and macaques (i4) were able to generalize in the
957 Nat condition, suggesting they were able to use the opponent motion cue in the task ($p=1.00$ mouse lemur,
958 $p=0.99$ macaque; ranksum test).
959
960

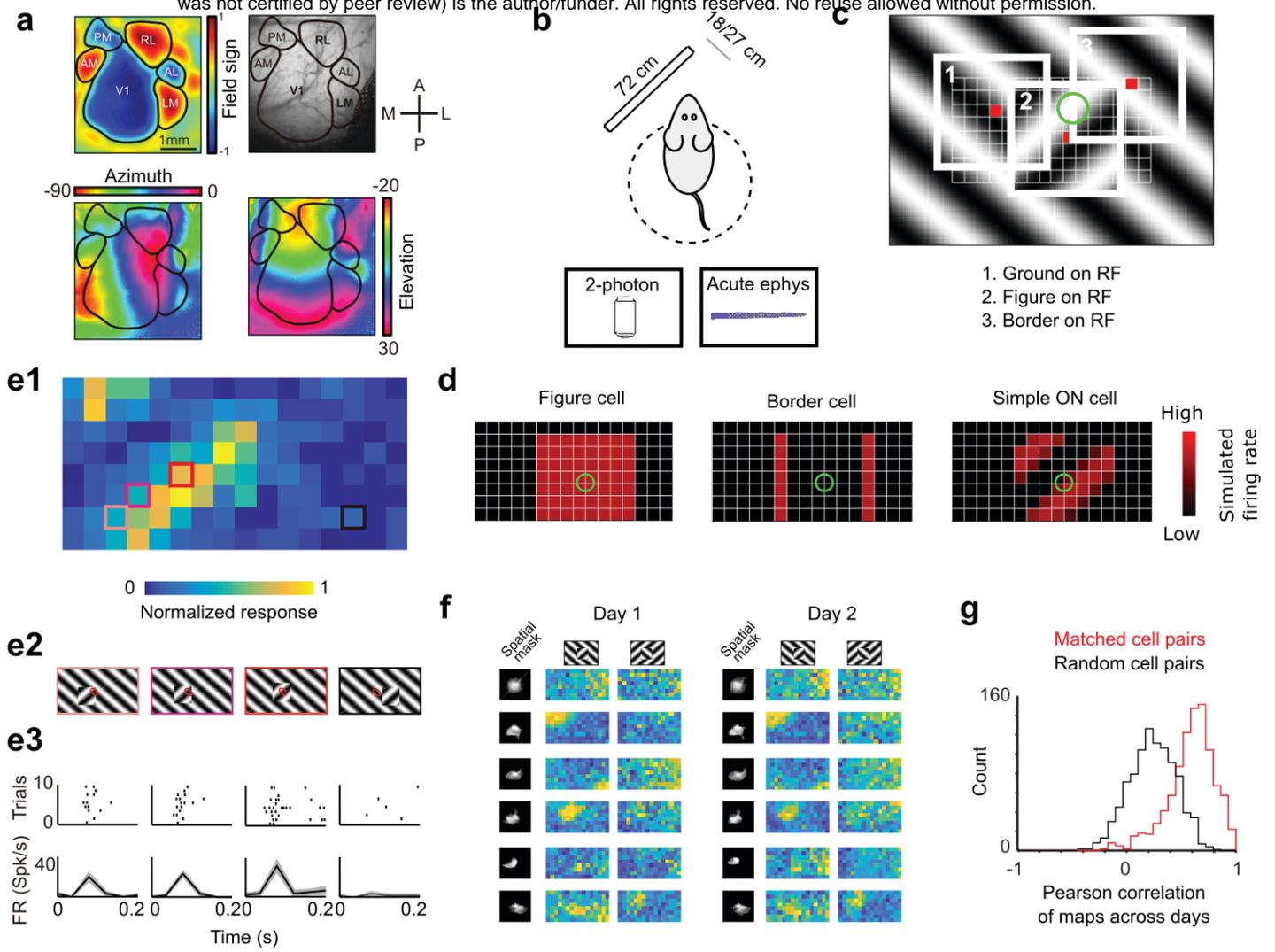


FIGURE 3

961 **Figure 3: Approach for measuring neural correlates of segmentation-related modulation in mouse**
962 **visual cortex.**

963
964 a) Widefield imaging of GCaMP transgenic animals was used to localize visual areas prior to neural
965 recording. A drifting bar stimulus was used to compute azimuth (bottom-left) and elevation (bottom-right)
966 maps across of the visual cortex. From these maps, a field-sign map (top-left) was computed, allowing
967 delineation of cortical visual areas (top-right). Alignment to vasculature maps guided subsequent
968 electrophysiology or 2-photon recordings to V1, LM, or RL.

969
970 b) Rodents were allowed to run freely on a spherical treadmill, with a 72 cm width (32 inch diagonal) screen
971 centered either 18 cm or 27 cm away from the left eye, while undergoing either electrophysiology or two-
972 photon imaging.

973
974 c) The stimulus consisted of a texture- and motion-defined square that was flashed randomly across a grid
975 of 16 horizontal positions x 8 vertical positions (128 positions total). On any given trial, a neuron with a
976 given receptive field location (schematized by the green circle) was stimulated by (1) ground, (2) figure, or
977 (3) edge, enabling us to measure both figure-ground and border-ownership modulation.

978
979 d) Schematic response maps to the stimulus in (c). Left: a “figure cell” responds only when a part of a figure
980 is over the receptive field. Middle: a “border cell” responds only when a figure border falls on the receptive
981 field and has orientation matching that of the cell (here assumed to be vertical). Right: a simple cell with an
982 ON subunit responds to the figure with phase dependence.

983
984 e1) Mean response at each of the 128 figure positions for an example V1 cell. Colored boxes correspond
985 to conditions shown in (e2).

986
987 e2) Four stimulus conditions outlined in (e1), ranging from receptive field on the figure (left) to receptive
988 field on the background (right).

989
990 e3) Raster (top) of spiking responses over 10 trials of each stimulus configuration and mean firing rate
991 (bottom). Error bars represent sem.

992
993 f) Example response maps from V1 using 2-photon calcium imaging show reliable responses from the same
994 neurons on successive days. Shown are 6 example neurons imaged across two days. Neurons were
995 matched according to a procedure described in the Methods. Colormap same as in (e1).

996
997 g) Distribution of Pearson correlations between figure maps for all matched cell pairs (red) and a set of
998 randomly-shuffled cell pairs (black). Neurons displayed highly reliable responses to the stimulus (N = 950
999 cell pairs in each group, mean=0.5579 for matched vs. mean=0.2054 for unmatched, $p = 1e-163$, ks-test).

1000
1001

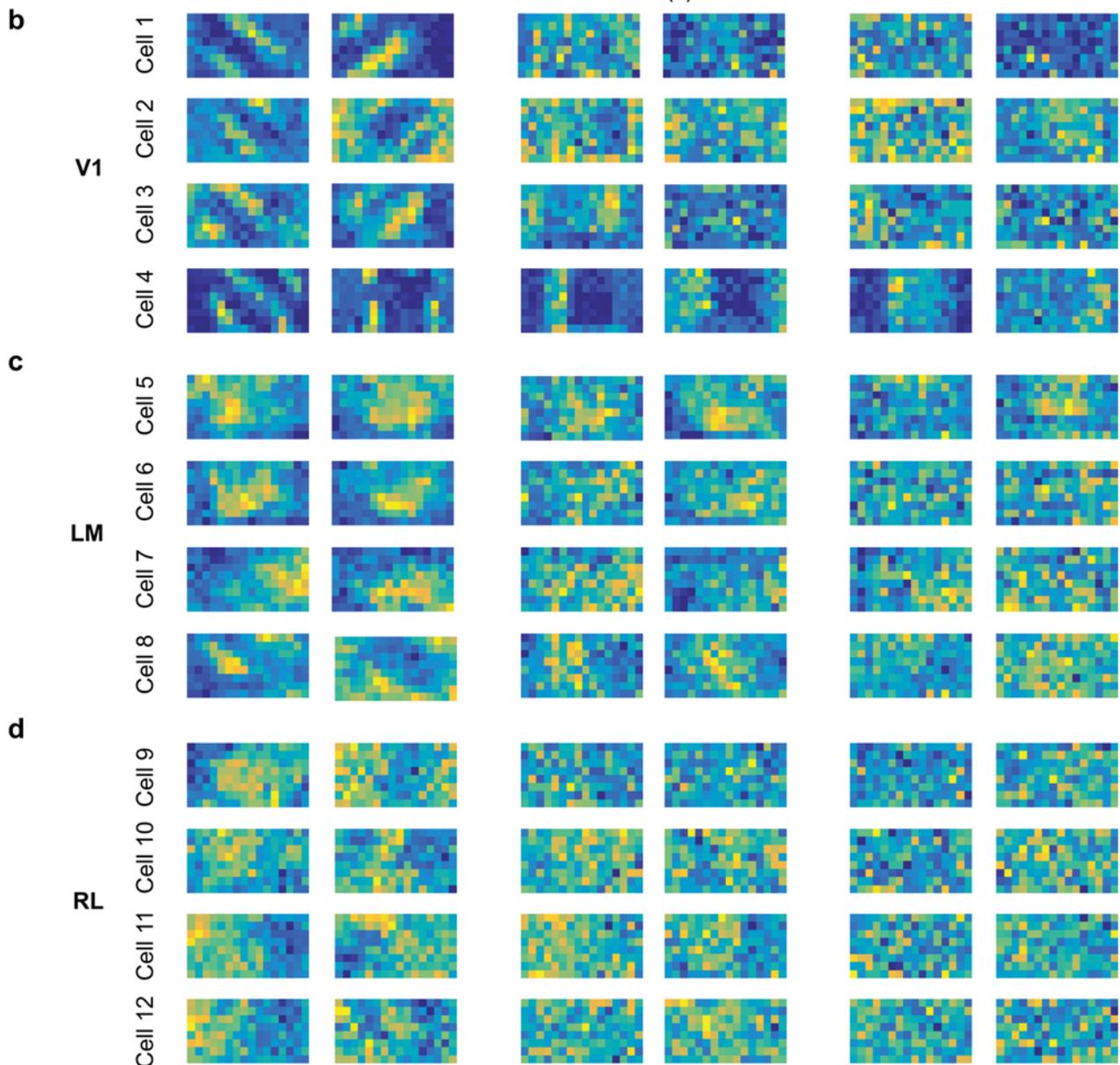
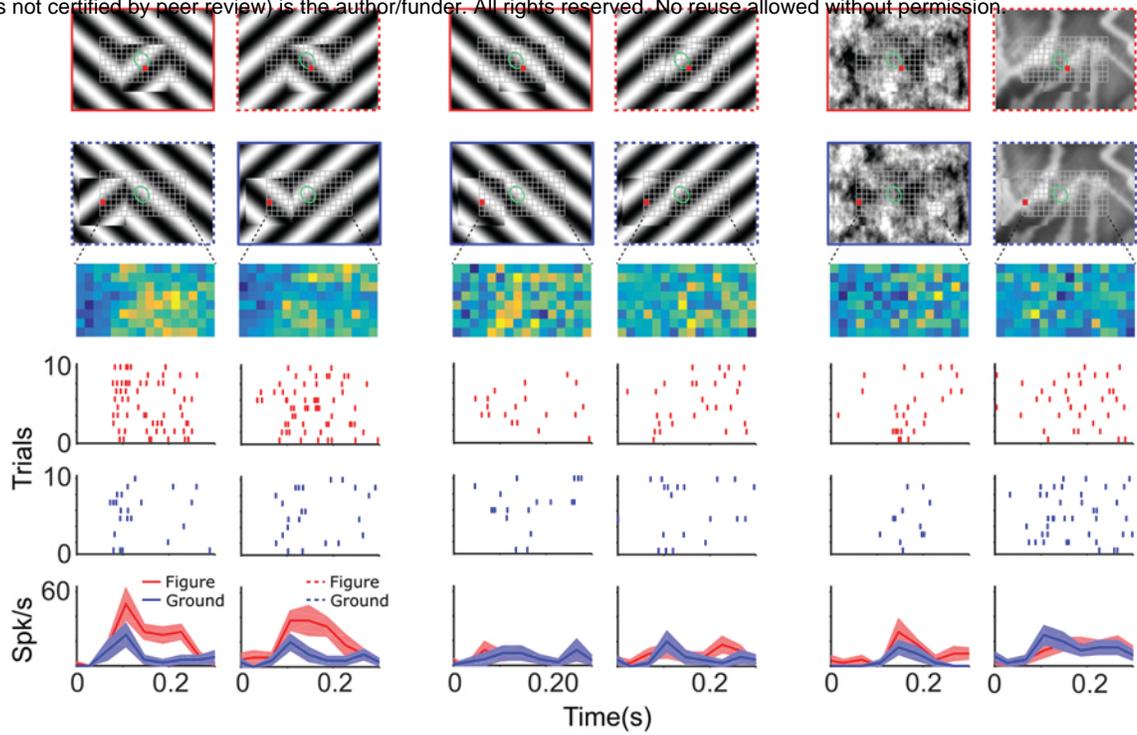


FIGURE 4

0  1
Normalized response

1002 **Figure 4: Segmentation-related modulation across mouse visual cortex is pattern dependent.**

1003

1004 a) To search for segmentation-related neural signals, we adapted the stimulus shown in **Fig. 3c** (a figure
1005 switching between 128 positions) to the three conditions that we had tested behaviorally (Cross, Iso, and
1006 Nat). As in the behavior experiments, for each condition we presented two variants (different orientations
1007 or patterns). Rows 1 & 2: two example frames (with square at different positions) are shown for each of the
1008 6 stimulus variants. Overlaid on these example stimuli are grids representing the 128 possible figure
1009 positions and a green ellipse representing the ON receptive field. Row 3: example figure map from one cell
1010 obtained for the conditions shown above. Rows 4 & 5: Example rasters when the figure was centered on
1011 (red) or off (blue) the receptive field. Row 6: PSTHs corresponding to the rasters; shaded error bars
1012 represent sem.

1013

1014 b) Figure maps for each of the 6 stimulus variants for four example neurons from V1 (responses measured
1015 using electrophysiology).

1016

1017 c) Figure maps for each of the 6 stimulus variants for four example neurons from LM.

1018

1019 d) Figure maps for each of the 6 stimulus variants for four example neurons from RL.

1020

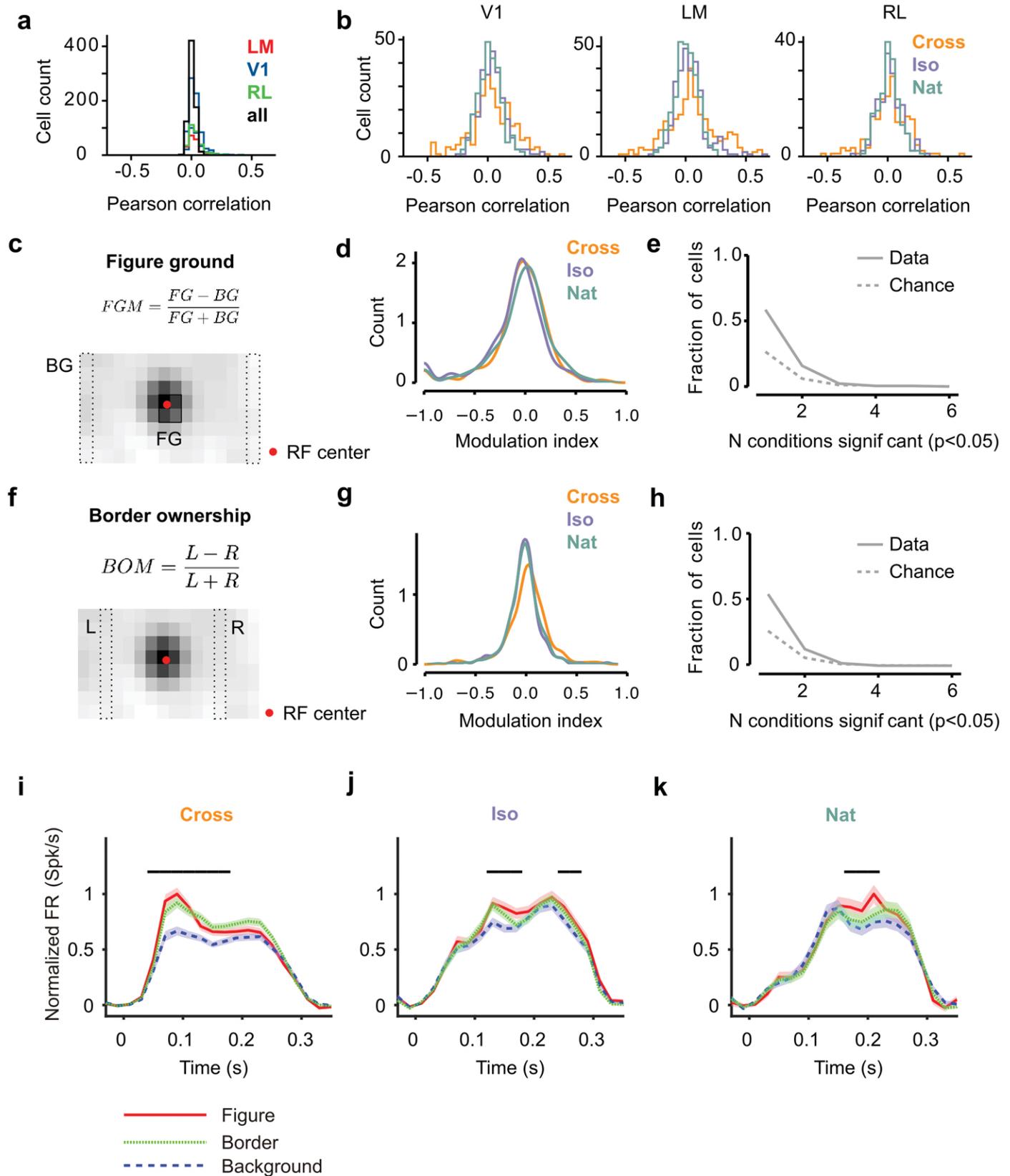


FIGURE 5

1021 **Figure 5: Mice lack consistent segmentation signals across texture conditions.**

1022 a) Distribution of Pearson correlations between figure maps across all $\binom{6}{2}$ pairs of conditions. No neuron
1023 in any area showed high correspondence (signified by non-zero mean) across all conditions tested,
1024 indicative of a texture-invariant figure response.

1025
1026 b) Distribution of Pearson correlations between figure maps across the two stimulus variants for each
1027 condition (orange: Cross, violet: Iso, green: Nat) and across visual areas (left: V1, middle: LM, right: RL).
1028

1029 c) A figure-ground modulation index was computed by taking the mean response on background trials
1030 (positions outlined by dashed lines) and the mean response on figure trials (positions outlined by solid line)
1031 and computing a normalized difference score.

1032
1033 d) Distribution (shown as a kernel density estimate) of figure ground modulation indices for Cross (orange),
1034 Iso (violet), and Nat (green) conditions, pooling cells from V1, LM, and RL.

1035
1036 e) Fraction of cells with figure-ground modulation index significantly different from zero ($p < 0.05$) for N
1037 stimulus variants (out of the six illustrated in **Fig. 4a**). Dotted gray line represents chance level false positive
1038 rate at $p < 0.05$ after 6 comparisons. For this analysis, FGM was computed similarly as (d), but responses
1039 were not averaged across orientations/patterns within each condition; thus each cell contributed 6 FGM
1040 values.

1041
1042 f) A border-ownership modulation index was computed by taking the mean response on left edge trials
1043 (positions outlined by dashed rectangle marked "L") and the mean response on right edge trials (positions
1044 outlined by dashed rectangle marked "R") and computing a normalized difference score.

1045
1046 g, h) Same as (d), (e), but for border-ownership modulation indices.

1047
1048 i, j, k) Mean response time courses across all cells in V1, RL, and LM to figure, ground, and border in the
1049 Cross (i), Iso (j), and Nat (k) conditions. Time points for which the response to the figure was significantly
1050 greater than the response to ground are indicated by the horizontal line above each plot ($p < 0.01$, t-test).
1051

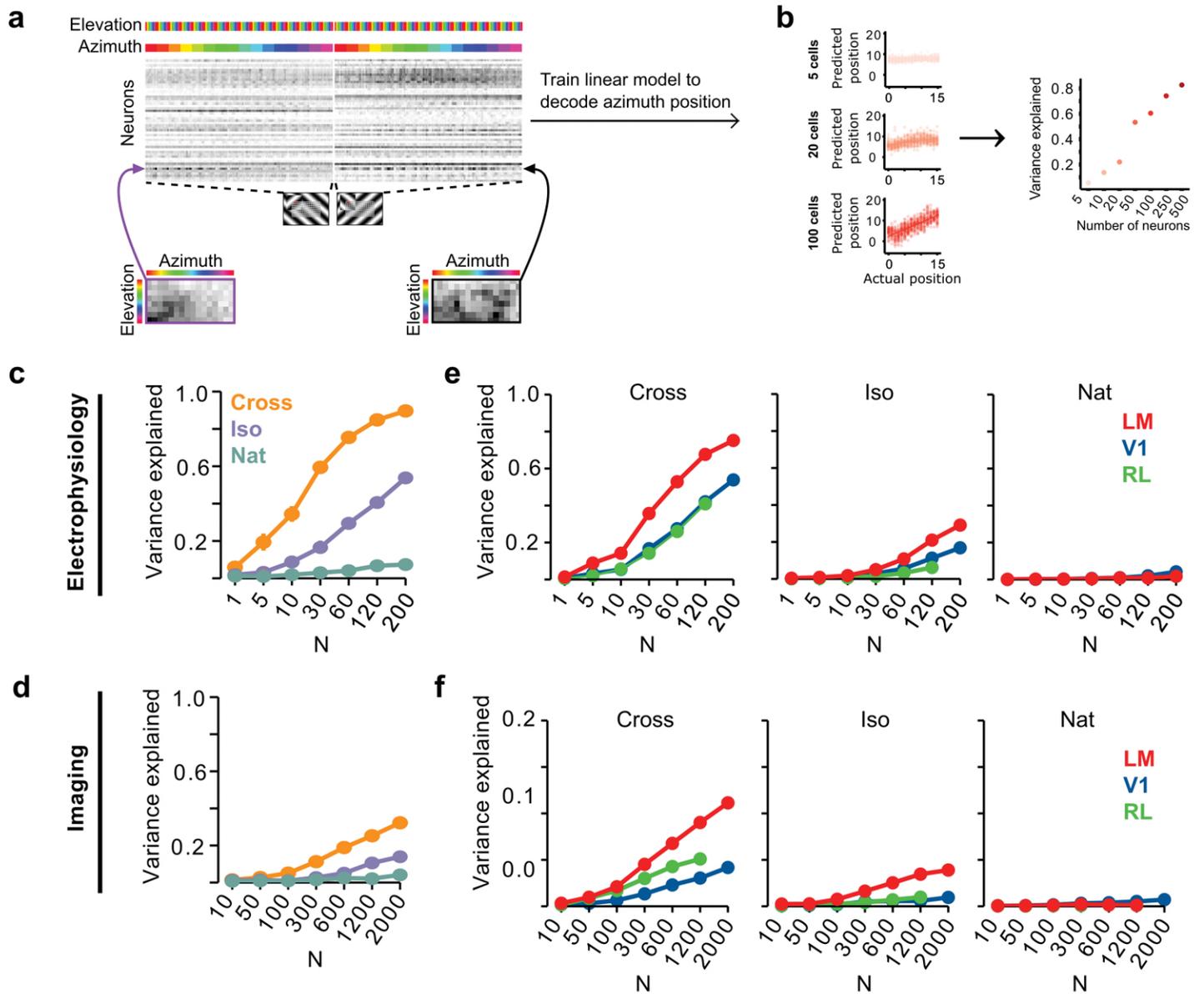


FIGURE 6

1052 **Figure 6: Decoding figure position from neural responses.**

1053

1054 a) Schematic of approach for decoding figure position from neural population responses. For each neuron,
1055 figure response maps for both types of stimuli from a given texture condition (Cross, Iso, Nat) were pooled,
1056 and reshaped into a 1-d vector, producing a population matrix of N neurons x N positions; the population
1057 response matrix for the Cross condition is shown. A linear decoder for figure azimuth position was then
1058 learned with cross validation from the population response matrix using 50% of trials for training and the
1059 remaining 50% of trials for testing.

1060

1061 b) A linear readout was computed for a given random sample of N neurons, shown here for 5 (top), 20
1062 (middle), and 100 (bottom) neurons. Each dot plots the actual azimuth bin (x-axis) against the predicted bin
1063 (y-axis). Mean explained variance was then computed across 50 repeated samples and plotted as a
1064 function of number of neurons (right).

1065

1066 c) Variance explained by decoded azimuth position as a function of number of neurons used to train the
1067 decoder for each of the different texture conditions (electrophysiology data). The most robust position
1068 decoding was obtained for Cross (orange), followed by Iso (violet) and then Nat (green). Error bars
1069 represent sem.

1070

1071 d) Same plot as (c) but for deconvolved calcium imaging data.

1072

1073 e) Same data as in (c), but broken down by both texture condition and visual region. LM (red) consistently
1074 showed better positional decoding than either V1 (blue) or RL (green). Error bars represent sem.

1075

1076 f) Same as (e) but for deconvolved calcium imaging data.

1077

1078

1079

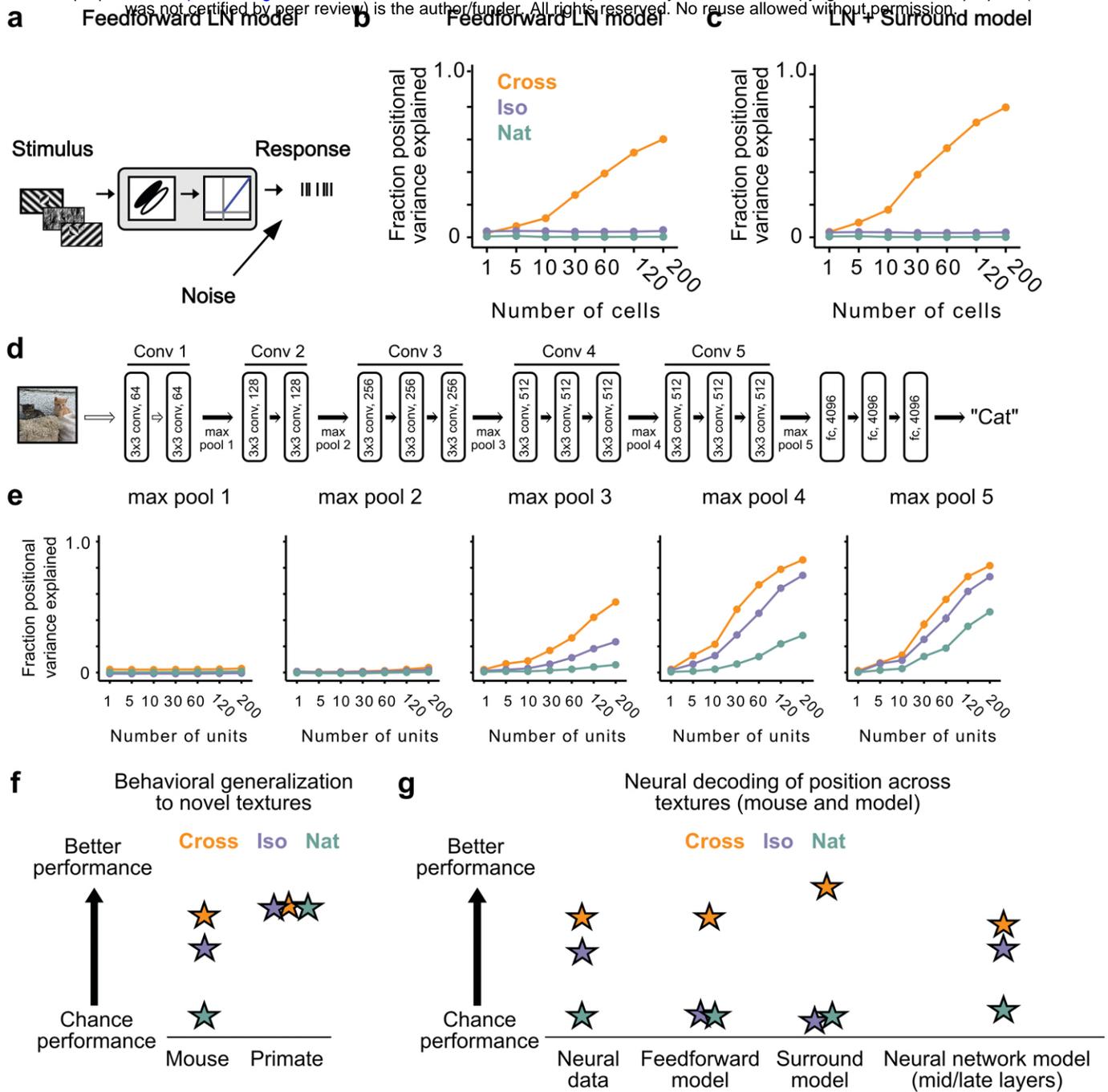


FIGURE 7

1080 **Figure 7: Mid to late layers of a deep network recapitulate mouse neural and behavioral performance**
1081 **on figure position decoding across texture conditions.**
1082

1083 a) Schematic of the feedforward linear-nonlinear (LN) encoding model (see Methods). The stimulus was
1084 passed through a Gabor filter, followed by a rectifying nonlinearity, and then a Poisson spiking model, with
1085 noise added to responses to simulate population response variability (e.g., due to non-sensory signals such
1086 as movement or arousal). We ran the same stimuli (128 positions x 6 conditions) through the model that
1087 we used for electrophysiology and two-photon imaging (cf. **Fig. 4a**).
1088

1089 b) Positional decoding performance, quantified as variance explained by decoded azimuth position, as a
1090 function of number of neurons in the feedforward LN model. Cross (orange) positional decoding was robust,
1091 while both Iso (violet) and Nat (green) were extremely poor, in contrast to electrophysiology (**Fig. 6c**) and
1092 imaging (**Fig. 6d**) results. Noise variance was set to twice the network-level firing rate within a condition
1093 here and in (c,e) below; for a full sweep across noise parameters, see **Extended Data Fig. 5**. Error bars
1094 represent sem. Small random offset added for visualization purposes.
1095

1096 c) Adding an orientation-dependent divisive term to the LN model to mimic iso-orientation surround
1097 suppression (**Extended Data Fig. 4**; LN + Surround model) yielded more robust decoding in the Cross
1098 condition (orange), but did not improve decoding in the Iso (violet) or Nat (green) conditions. For a full
1099 sweep across noise parameters, see **Extended Data Fig. 5**. Error bars represent sem. Small random offset
1100 added for visualization purposes.
1101

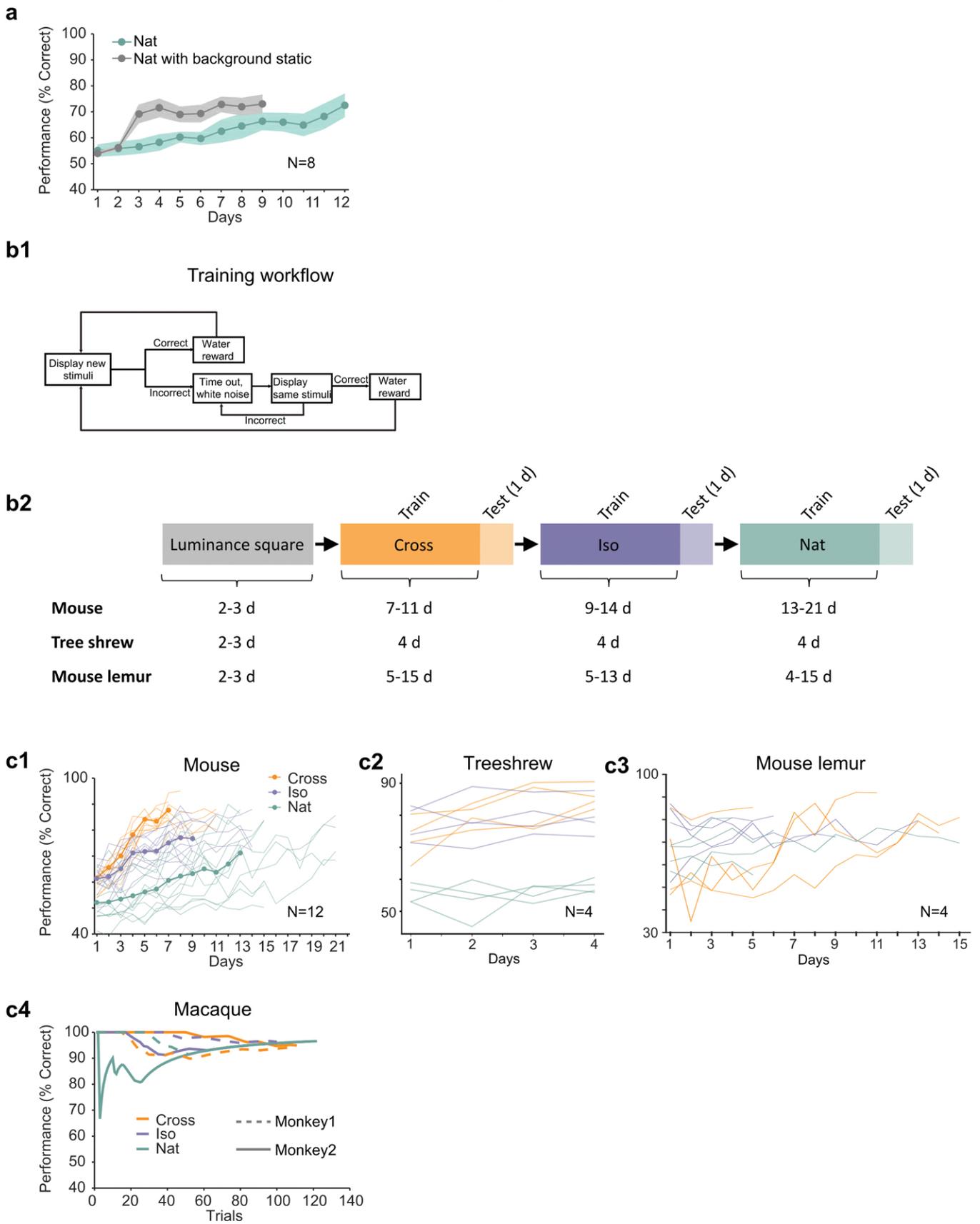
1102 d) Architecture of a pre-trained deep neural network (VGG16) trained on image recognition [36]. Five
1103 convolution layers are followed by three fully connected layers.
1104

1105 e) Positional decoding performance increases throughout the network with most robust decoding in layer
1106 4. In mid to late layers (3-5) of the deep network, decoding performance was best for Cross (orange),
1107 followed by Iso (violet) and then Nat (green), mirroring mouse behavioral performance (**Fig. 2c, i1**) and
1108 neural data (**Fig. 6c, d**). For a full sweep across noise parameters, see **Extended Data Fig. 6**. Error bars
1109 represent sem.
1110

1111 f) Schematic summary of behavioral results. Mice showed texture-dependent performance in a figure
1112 localization task, with Cross > Iso > Nat, despite the presence of a common motion cue for segmentation
1113 in all conditions. In contrast, primates showed no dependence on carrier texture and were able to use the
1114 differential motion cue to perform the task.
1115

1116 g) Schematic summary of neural and modeling results, Positional decoding from neural populations in V1,
1117 LM, and RL mirror the textural dependence of the behavior, with Cross > Iso > Nat. This ordering in
1118 performance was not captured by a feedforward LN model or an LN model with surround interactions.
1119 However, it emerged naturally from non-linear interactions in mid to late layers of a deep neural network.
1120

1121
1122
1123



Extended Data Fig.1

1124 **Extended Data Figure 1: Additional behavioral performance statistics.**

1125

1126 a) Averaged learning curves for Nat condition (n=12) and Nat condition with static background (n=6). If the
1127 task is turned into a pure local motion detection task by making the background static, mice learn
1128 considerably faster, demonstrating that they are able to detect the motion in the Nat condition. Error bars
1129 indicate sem.

1130

1131 b1) Schematic of training procedure indicating flow of stimuli within a training session.

1132 b2) Schematic of training procedure used for **Fig. 2i1-3**. Animals were first trained on a luminance square.
1133 Then they were trained on Cross, followed by Iso and then Nat. For each condition, the training period was
1134 followed by one day of generalization testing.

1135

1136 c1) Individual learning curves of 12 mice for Cross, Iso, and Nat conditions.

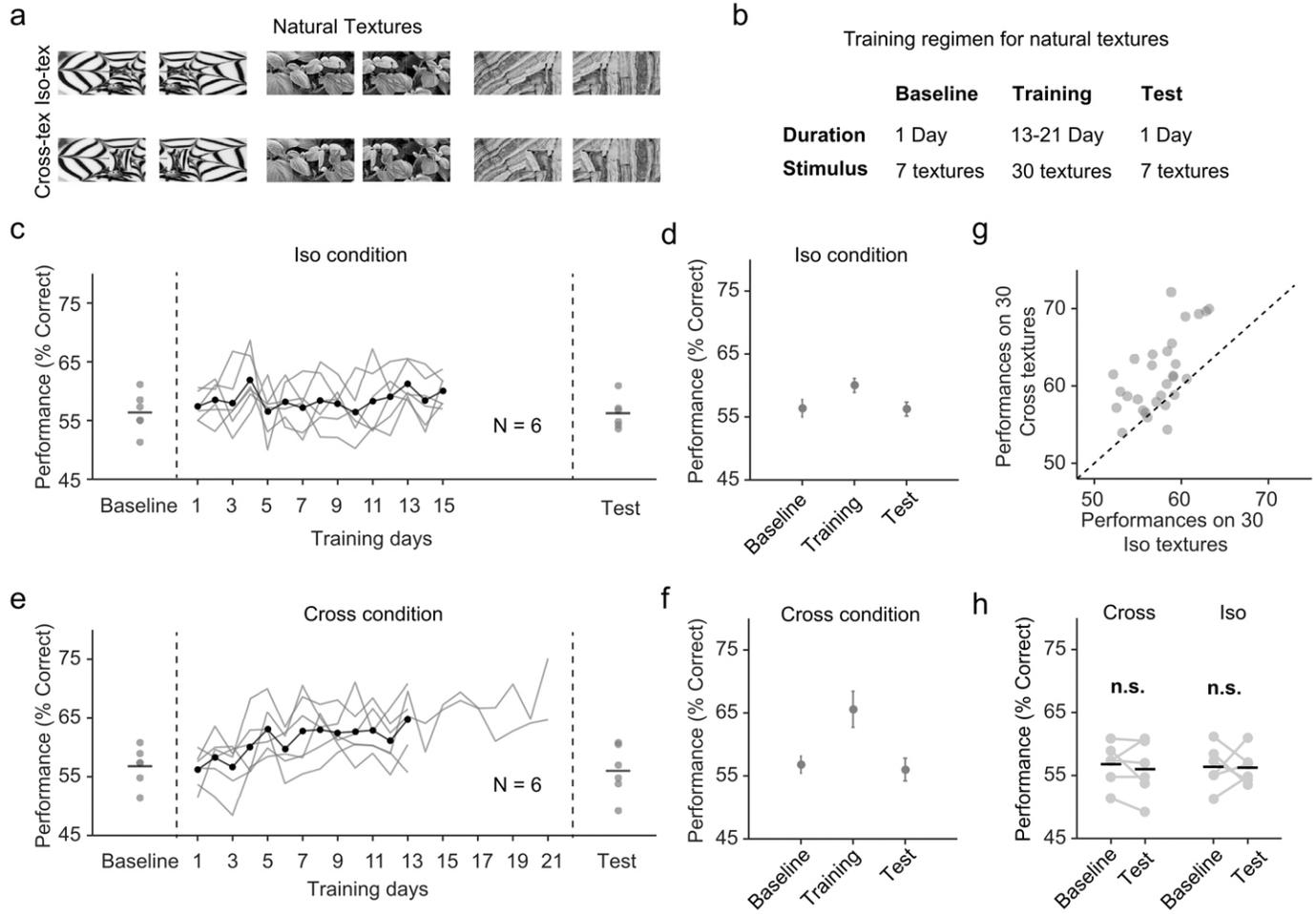
1137 c2) Individual learning curves of 4 treeshrews for Cross, Iso, and Nat conditions. Animals could already
1138 perform Cross and Iso > 70% on the first training session, displaying much faster learning than mice.

1139 c3) Individual learning curves of 4 mouse lemurs for Cross, Iso, and Nat conditions.

1140 c4) Learning curves of two monkeys for Cross, Iso, and Nat conditions (monkey A: solid lines, monkey B:
1141 dashed lines). Both animals rapidly learned within a single session to perform the task (after only a brief
1142 period of training with a luminance-defined square at the beginning of the session).

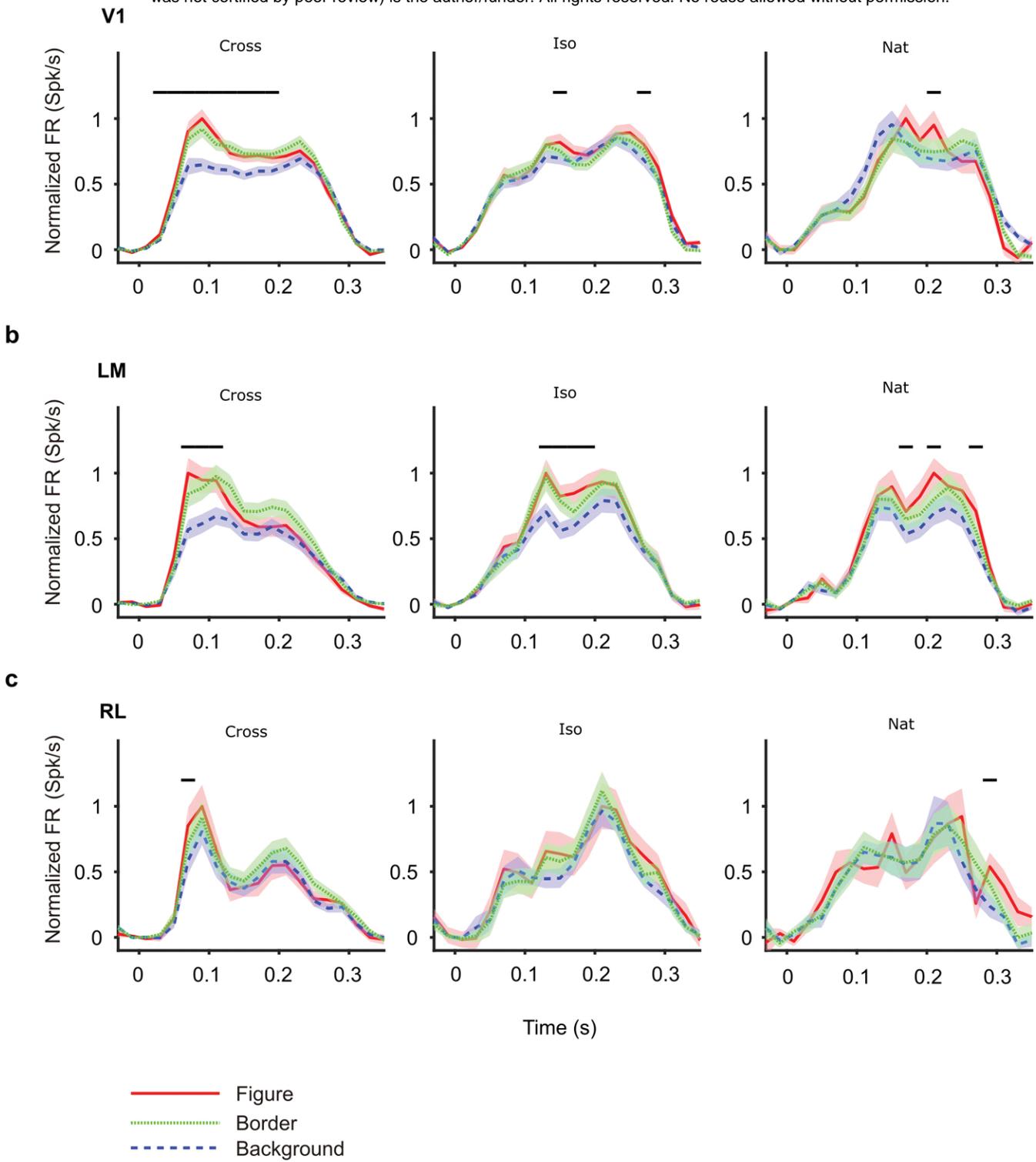
1143

1144

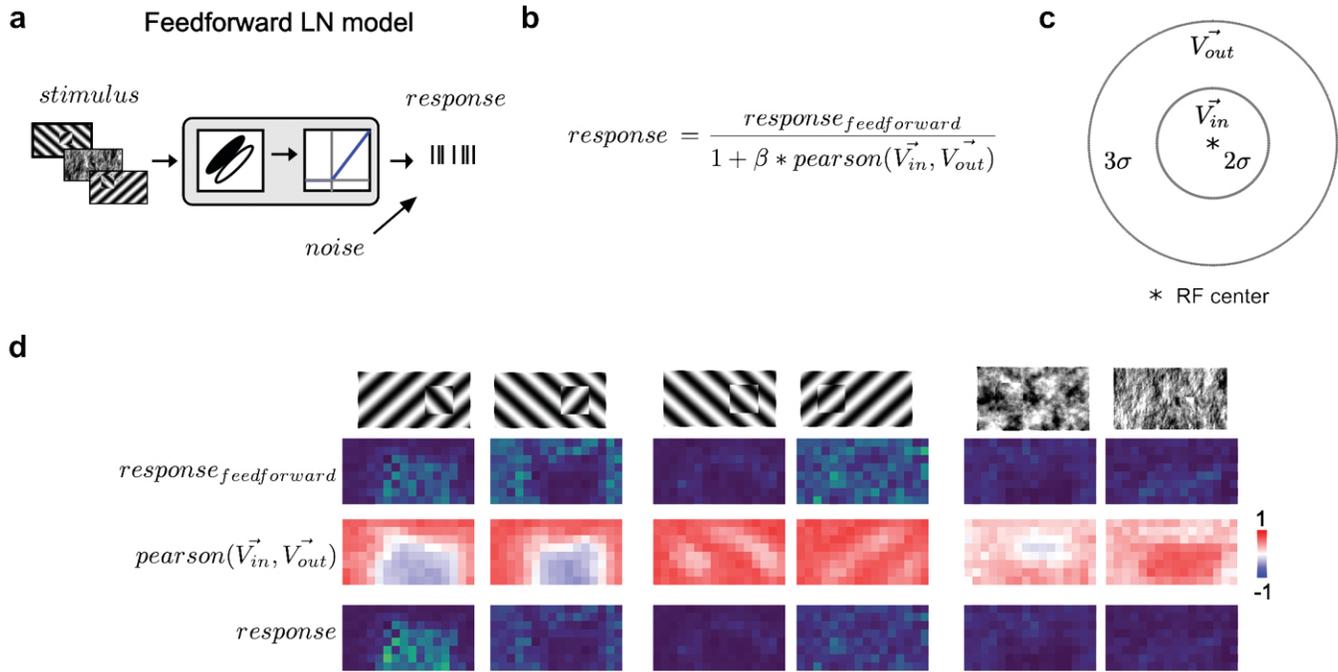


Extended Data Fig. 2

- 1145 **Extended Data Figure 2: Natural texture task shows advantage for learning figures with cross-**
1146 **oriented energy.**
1147
1148 a) Examples of natural textures in Cross and Iso conditions.
1149
1150 b) Table summarizing the training paradigm for the natural texture task.
1151
1152 c) Performance on baseline day, training days, and test day in Iso condition. Solid black line indicates the
1153 mean of training curves on training days (N=10).
1154
1155 d) Mean performance on baseline day, training days, and test day in Iso condition. Mice showed no
1156 improvement in performance on novel textures after showing behavioral increases during the training
1157 period. Error bars represent sem.
1158
1159 e) Same as (c) but for Cross condition.
1160
1161 f) Same as (d) but for Cross condition. Just as before, mice showed no improvement in performance on
1162 novel textures after showing behavioral increases during the training period.
1163
1164 g) Comparing performance between Cross and Iso conditions on training days. Each point represents the
1165 mean performance of a given texture on both Cross and Iso conditions. Most points lie above the unity line
1166 ($p = 5.9 \times 10^{-5}$, sign test).
1167
1168 h) Performance for Cross and Iso conditions on baseline and test day (Cross, $p = 0.3$; Iso, $p = 0.97$; rank
1169 sum test).
1170



1171 **Extended Data Figure 3: Mean time courses of responses across the population to figure, ground,**
1172 **and border in areas V1, LM, and RL. Conventions as in Fig. 4.**



1173 **Extended Data Figure 4: Modeling orientation-dependent surround interactions.**

1174

1175 a) Standard feedforward LN model used to model simple cell responses.

1176

1177 b) Model for divisive normalization that we used to model orientation dependent surround interactions [66].
1178 The neuron's feedforward response was modulated by a divisive term that took into account the mean
1179 response of all neurons of a given Gabor type with RF centers within 2σ of a given receptive field's center
1180 (\vec{V}_{in}) and those that lay $> 2\sigma$ and $< 5\sigma$ outside of the receptive field (\vec{V}_{out}). From these two vectors (each
1181 100 elements long), we computed a Pearson correlation, $pearson(\vec{V}_{in}, \vec{V}_{out})$, which ranged from -1 to 1,
1182 leading to suppression when the orientation energy in \vec{V}_{in} and \vec{V}_{out} matched and facilitation when \vec{V}_{in} and
1183 \vec{V}_{out} were orthogonal.

1184

1185 c) Schematic representation of the zones for \vec{V}_{in} and \vec{V}_{out} , each defined relative to a cell's receptive field
1186 center.

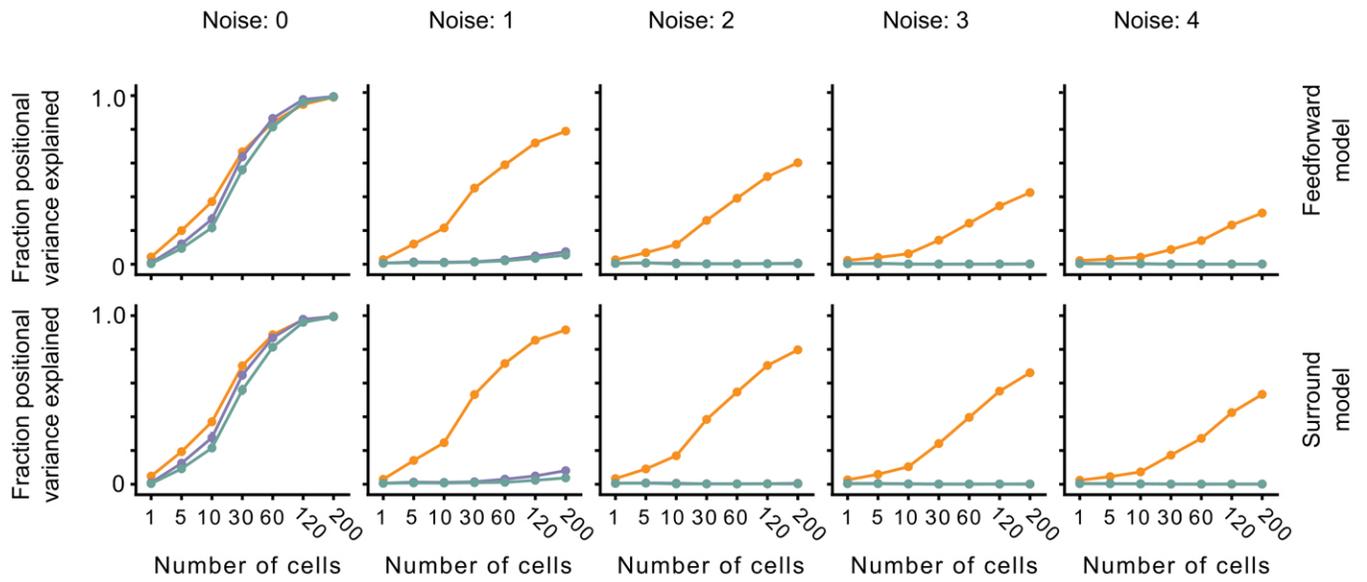
1187

1188 d) Row 1: Schematic of six stimulus conditions. Row 2: Figure maps for an example feedforward LN neuron
1189 for each of the 6 conditions. Row 3: Map of the modulation term value at each position; note that whether
1190 there is suppression or facilitation is a function of the stimulus condition, figure position, and receptive field
1191 position. Row 4: Figure maps for an example surround-model neuron; for the Cross condition, adding
1192 surround modulation results in general facilitation, while for Iso and Nat conditions, it results in depression
1193 or no modulation.

1194

1195

1196

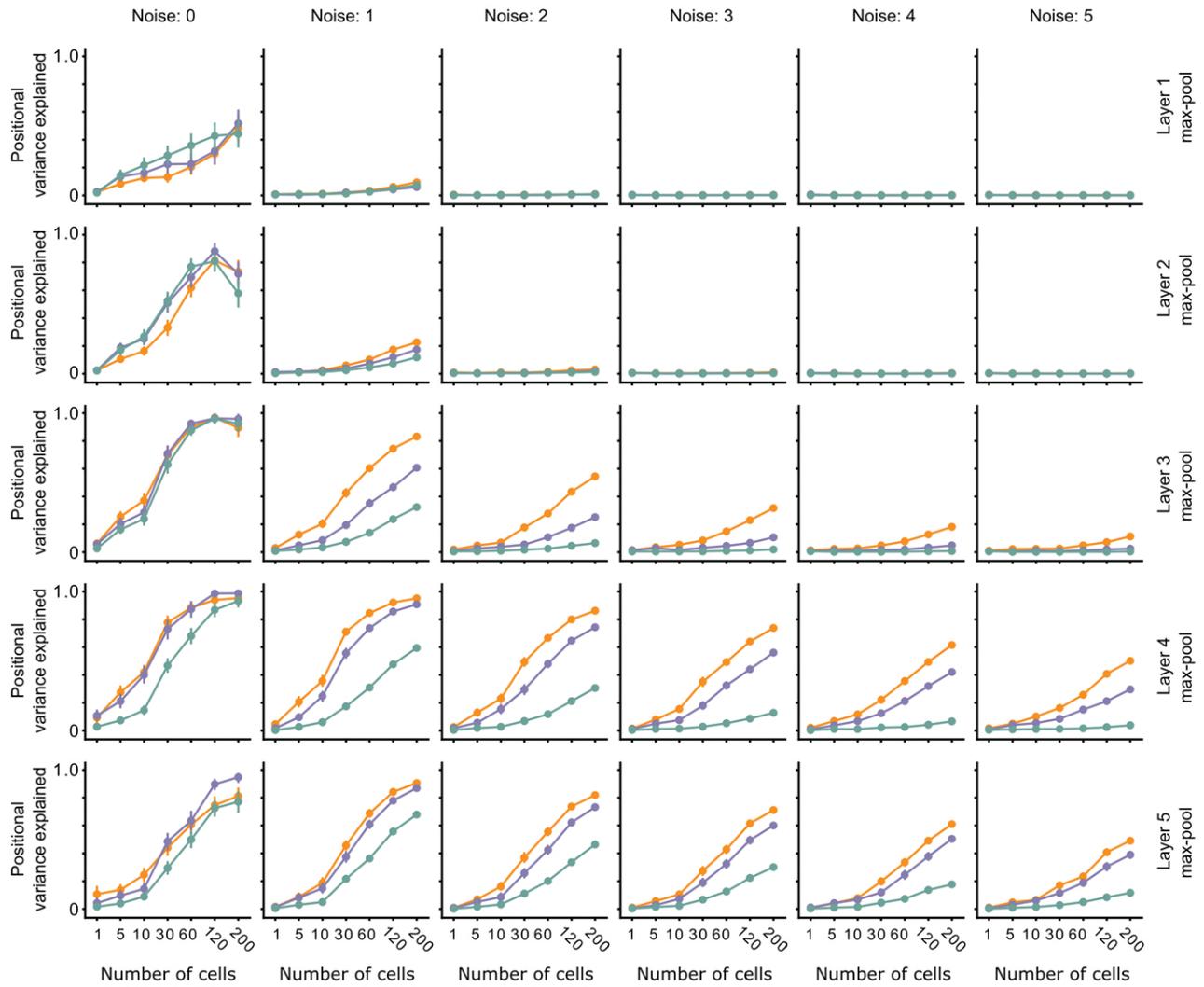


Extended Data Fig. 5

1197 **Extended Data Figure 5: The effect of noise on position decoding for feedforward LN and surround**
1198 **models.**

1199
1200 Top row: Position decoding performance for Cross, Iso, and Nat conditions in a population of feedforward
1201 LN neurons across noise conditions (columns). Note that positional decoding for Cross is high across noise
1202 conditions, while for Iso and Nat it remains low. The noise levels indicate the ratio between the noise
1203 variance and the network-level firing rate (see Methods).

1204
1205 Bottom row: Same as top row, for a population of feedforward LN neurons with orientation-dependent
1206 surround interactions. Including this extra-classical receptive field modulation had no effect on the relative
1207 decoding performance for Iso versus Nat conditions. Error bars represent sem.
1208
1209



Extended Data Fig. 6

1210 **Extended Data Figure 6: The effect of noise on position decoding for intermediate layers of VGG16.**
1211
1212 Position decoding performance for Cross, Iso, and Nat conditions as a function of neural network layer
1213 (rows) and increasing population response noise (columns). Note that across various noise conditions, the
1214 separation of Cross, Iso, and Nat conditions remains prominent. Error bars represent sem.
1215
1216

1217 **References**

- 1218
- 1219
- 1220 1. Nakayama, K, He, ZJ, and Shimojo, S, *Visual surface representation: A critical link between lower-*
- 1221 *level and higher-level vision*, in **An invitation to cognitive science**, S.M. Kosslyn and D.N.
- 1222 Osherson, Editors. **1995**, MIT Press. p. 1–70.
- 1223 2. Lamme, VA, *The neurophysiology of figure-ground segregation in primary visual cortex*. **The**
- 1224 **Journal of Neuroscience: The Official Journal of the Society for Neuroscience**, **1995**. 15(2):
- 1225 p. 1605-1615.
- 1226 3. Poort, J, Raudies, F, Wannig, A, Lamme, VAF, Neumann, H, and Roelfsema, PR, *The role of*
- 1227 *attention in figure-ground segregation in areas V1 and V4 of the visual cortex*. **Neuron**, **2012**. 75(1):
- 1228 p. 143-156.
- 1229 4. Qiu, FT and Heydt, Rvd, *Figure and Ground in the Visual Cortex: V2 Combines Stereoscopic Cues*
- 1230 *with Gestalt Rules*. **Neuron**, **2005**. 47(1): p. 155-166.
- 1231 5. Evans, DA, Stempel, AV, Vale, R, Ruehle, S, Lefler, Y, and Branco, T, *A synaptic threshold*
- 1232 *mechanism for computing escape decisions*. **Nature**, **2018**. 558(7711): p. 590-594.
- 1233 6. Fiser, A, Mahringer, D, Oyibo, HK, Petersen, AV, Leinweber, M, and Keller, GB, *Experience-*
- 1234 *dependent spatial expectations in mouse visual cortex*. **Nature Neuroscience**, **2016**. 19(12): p.
- 1235 1658-1664.
- 1236 7. Harvey, CD, Coen, P, and Tank, DW, *Choice-specific sequences in parietal cortex during a virtual-*
- 1237 *navigation decision task*. **Nature**, **2012**. 484(7392): p. 62-68.
- 1238 8. Hoy, JL, Yavorska, I, Wehr, M, and Niell, CM, *Vision Drives Accurate Approach Behavior during*
- 1239 *Prey Capture in Laboratory Mice*. **Current biology: CB**, **2016**. 26(22): p. 3046-3052.
- 1240 9. Leinweber, M, Ward, DR, Sobczak, JM, Attinger, A, and Keller, GB, *A Sensorimotor Circuit in*
- 1241 *Mouse Cortex for Visual Flow Predictions*. **Neuron**, **2017**. 96(5): p. 1204.
- 1242 10. Pinto, L, Goard, MJ, Estandian, D, Xu, M, Kwan, AC, Lee, S-H, Harrison, TC, Feng, G, and Dan,
- 1243 Y, *Fast modulation of visual perception by basal forebrain cholinergic neurons*. **Nature**
- 1244 **Neuroscience**, **2013**. 16(12): p. 1857-1863.
- 1245 11. Abbott, LF, Angelaki, DE, Carandini, M, Churchland, AK, Dan, Y, Dayan, P, Deneve, S, Fiete, I,
- 1246 Ganguli, S, Harris, KD, Häusser, M, Hofer, S, Latham, PE, Mainen, ZF, Mrsic-Flogel, T, Paninski,
- 1247 L, Pillow, JW, Pouget, A, Svoboda, K, Witten, IB, and Zador, AM, *An International Laboratory for*
- 1248 *Systems and Computational Neuroscience*. **Neuron**, **2017**. 96(6): p. 1213-1218.
- 1249 12. Pinto, L, Rajan, K, DePasquale, B, Thiberge, SY, Tank, DW, and Brody, CD, *Task-Dependent*
- 1250 *Changes in the Large-Scale Dynamics and Necessity of Cortical Regions*. **Neuron**, **2019**. 104(4):
- 1251 p. 810-824 e9.
- 1252 13. Marshel, James H, Garrett, Marina E, Nauhaus, I, and Callaway, Edward M, *Functional*
- 1253 *Specialization of Seven Mouse Visual Cortical Areas*. **Neuron**, **2011**. 72(6): p. 1040-1054.
- 1254 14. Niell, CM and Stryker, MP, *Highly Selective Receptive Fields in Mouse Visual Cortex*. **Journal of**
- 1255 **Neuroscience**, **2008**. 28(30): p. 7520-7536.
- 1256 15. Self, MW, Lorteije, JAM, Vangeneugden, J, van Beest, EH, Grigore, ME, Levelt, CN, Heimel, JA,
- 1257 and Roelfsema, PR, *Orientation-tuned surround suppression in mouse visual cortex*. **The Journal**
- 1258 **of Neuroscience: The Official Journal of the Society for Neuroscience**, **2014**. 34(28): p. 9290-
- 1259 9304.
- 1260 16. Keller, AJ, Dipoppa, M, Roth, MM, Caudill, MS, Ingrosso, A, Miller, KD, and Scanziani, M, *A*
- 1261 *Disinhibitory Circuit for Contextual Modulation in Primary Visual Cortex*. **Neuron**, **2020**. 108(6): p.
- 1262 1181-1193 e8.
- 1263 17. Frost, BJ and Nakayama, K, *Single visual neurons code opposing motion independent of direction*.
- 1264 **Science (New York, N.Y.)**, **1983**. 220(4598): p. 744-745.
- 1265 18. He, ZJ and Nakayama, K, *Surfaces versus features in visual search*. **Nature**, **1992**. 359(6392): p.
- 1266 231-233.
- 1267 19. Williford, JR and von der Heydt, R, *Border-ownership coding*. **Scholarpedia journal**, **2013**. 8(10):
- 1268 p. 30040.
- 1269 20. Zhou, H, Friedman, HS, and von der Heydt, R, *Coding of border ownership in monkey visual cortex*.
- 1270 **The Journal of Neuroscience: The Official Journal of the Society for Neuroscience**, **2000**.
- 1271 20(17): p. 6594-6611.
- 1272 21. Tsao, T and Tsao, DY *A topological solution to object segmentation and tracking*. **arXiv**, **2021**.

- 1273 22. Hall, JR, Cuthill, IC, Baddeley, R, Shohet, AJ, and Scott-Samuel, NE, *Camouflage, detection and*
1274 *identification of moving targets. Proceedings of the Royal Society B: Biological Sciences, 2013.*
1275 280(1758).
- 1276 23. Gibson, JJ, *The Ecological Approach to Visual Perception.* 1979: Houghton Mifflin.
- 1277 24. Nakayama, K and Shimojo, S, *da Vinci stereopsis: depth and subjective occluding contours from*
1278 *unpaired image points. Vision Research, 1990.* 30(11): p. 1811-1825.
- 1279 25. McCulloch, W and Pitts, W, *On how we know universals: The perception of auditory and visual*
1280 *forms. Bulletin of Mathematical Biophysics, 1947:* p. 127-147.
- 1281 26. DiCarlo, JJ and Cox, DD, *Untangling invariant object recognition. Trends Cogn Sci, 2007.* 11(8):
1282 p. 333-41.
- 1283 27. Rossi, AF, Desimone, R, and Ungerleider, LG, *Contextual modulation in primary visual cortex of*
1284 *macaques. The Journal of Neuroscience: The Official Journal of the Society for*
1285 *Neuroscience, 2001.* 21(5): p. 1698-1709.
- 1286 28. Hesse, JK and Tsao, DY, *Consistency of Border-Ownership Cells across Artificial Stimuli, Natural*
1287 *Stimuli, and Stimuli with Ambiguous Contours. The Journal of Neuroscience: The Official*
1288 *Journal of the Society for Neuroscience, 2016.* 36(44): p. 11338-11349.
- 1289 29. Kirchberger, L, Mukherjee, S, Schnabel, UH, van Beest, EH, Barsegyan, A, Levelt, CN, Heimel,
1290 JA, Lorteije, JAM, van der Togt, C, Self, MW, and Roelfsema, PR, *The essential role of feedback*
1291 *processing for figure-ground perception in mice. bioRxiv, 2020:* p. 456459.
- 1292 30. Schnabel, UH, Bossens, C, Lorteije, JAM, Self, MW, Beeck, HOd, and Roelfsema, PR, *Figure-*
1293 *ground perception in the awake mouse and neuronal activity elicited by figure-ground stimuli in*
1294 *primary visual cortex. Scientific Reports, 2018.* 8(1): p. 1-14.
- 1295 31. Keller, AJ, Roth, MM, and Scanziani, M, *Feedback generates a second receptive field in neurons*
1296 *of the visual cortex. Nature, 2020.* 582(7813): p. 545-549.
- 1297 32. Zipser, K, Lamme, VA, and Schiller, PH, *Contextual modulation in primary visual cortex. J*
1298 *Neurosci, 1996.* 16(22): p. 7376-89.
- 1299 33. Wang, Q, Gao, E, and Burkhalter, A, *Gateways of ventral and dorsal streams in mouse visual*
1300 *cortex. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience,*
1301 *2011.* 31(5): p. 1905-1918.
- 1302 34. Berens, P, Freeman, J, Deneux, T, Chenkov, N, McColgan, T, Speiser, A, Macke, JH, Turaga, SC,
1303 Mineault, P, Rupprecht, P, Gerhard, S, Friedrich, RW, Friedrich, J, Paninski, L, Pachitariu, M,
1304 Harris, KD, Bolte, B, Machado, TA, Ringach, D, Stone, J, Rogerson, LE, Sofroniew, NJ, Reimer, J,
1305 Froudarakis, E, Euler, T, Rosón, MR, Theis, L, Tolias, AS, and Bethge, M, *Community-based*
1306 *benchmarking improves spike rate inference from two-photon calcium imaging data. PLOS*
1307 *Computational Biology, 2018.* 14(5): p. e1006157.
- 1308 35. Vinken, K and Op de Beeck, H, *Using deep neural networks to evaluate object vision tasks in rats.*
1309 *PLOS Computational Biology, 2021.*
- 1310 36. Simonyan, K and Zisserman, A, *Very Deep Convolutional Networks for Large-Scale Image*
1311 *Recognition. arXiv:1409.1556 [cs], 2014.*
- 1312 37. Ho, CLA, Zimmermann, R, Florez Weidinger, JD, Prsa, M, Schottdorf, M, Merlin, S, Okamoto, T,
1313 Ikezoe, K, Pifferi, F, Aujard, F, Angelucci, A, Wolf, F, and Huber, D, *Orientation Preference Maps*
1314 *in Microcebus murinus Reveal Size-Invariant Design Principles in Primate Visual Cortex. Curr Biol,*
1315 *2021.* 31(4): p. 733-741 e7.
- 1316 38. Wong, P and Kaas, JH, *Architectonic subdivisions of neocortex in the tree shrew (Tupaia*
1317 *belangeri). Anat Rec (Hoboken), 2009.* 292(7): p. 994-1027.
- 1318 39. Mustafar, F, Harvey, MA, Khani, A, Arato, J, and Rainer, G, *Divergent Solutions to Visual Problem*
1319 *Solving across Mammalian Species. eNeuro, 2018.* 5(4).
- 1320 40. Van Hooser, SD, Roy, A, Rhodes, HJ, Culp, JH, and Fitzpatrick, D, *Transformation of receptive*
1321 *field properties from lateral geniculate nucleus to superficial V1 in the tree shrew. J Neurosci,*
1322 *2013.* 33(28): p. 11494-505.
- 1323 41. Roe, AW, Chelazzi, L, Connor, CE, Conway, BR, Fujita, I, Gallant, JL, Lu, H, and Vanduffel, W,
1324 *Toward a unified theory of visual area V4. Neuron, 2012.* 74(1): p. 12-29.
- 1325 42. Keyser, RD, Bossens, C, Kubilius, J, and Beeck, HPOd, *Cue-invariant shape recognition in rats as*
1326 *tested with second-order contours. Journal of Vision, 2015.* 15(15): p. 14-14.
- 1327 43. Schnell, AE, Van den Bergh, G, Vermaercke, B, Gijbels, K, Bossens, C, and de Beeck, HO, *Face*
1328 *categorization and behavioral templates in rats. Journal of Vision, 2019.* 19(14): p. 9.

- 1329 44. Vermaercke, B and Op de Beeck, HP, *A Multivariate Approach Reveals the Behavioral Templates*
1330 *Underlying Visual Discrimination in Rats*. **Current Biology**, 2012. 22(1): p. 50-55.
- 1331 45. Yu, Y, Hira, R, Stirman, JN, Yu, W, Smith, IT, and Smith, SL, *Mice use robust and common*
1332 *strategies to discriminate natural scenes*. **Scientific Reports**, 2018. 8(1): p. 1-13.
- 1333 46. Zoccolan, D, *Invariant visual object recognition and shape processing in rats*. **Behavioural Brain**
1334 **Research**, 2015. 285: p. 10-33.
- 1335 47. Zoccolan, D, Oertelt, N, DiCarlo, JJ, and Cox, DD, *A rodent model for the study of invariant visual*
1336 *object recognition*. **Proceedings of the National Academy of Sciences of the United States of**
1337 **America**, 2009. 106(21): p. 8748-8753.
- 1338 48. Schnabel, UH, Kirchberger, L, Beest, EHv, Mukherjee, S, Barsegyan, A, Lorteije, JAM, Togt, Cvd,
1339 Self, MW, and Roelfsema, PR, *Feedforward and feedback processing during figure-ground*
1340 *perception in mice*. **bioRxiv**, 2018: p. 456459.
- 1341 49. Kim, T, Soto, F, and Kerschensteiner, D, *An excitatory amacrine cell detects object motion and*
1342 *provides feature-selective input to ganglion cells in the mouse retina*. **eLife**, 2015. 4.
- 1343 50. Ölveczky, BP, Baccus, SA, and Meister, M, *Segregation of object and background motion in the*
1344 *retina*. **Nature**, 2003. 423(6938): p. 401-408.
- 1345 51. Goodfellow, IJ, Shlens, J, and Szegedy, C, *Explaining and Harnessing Adversarial Examples*.
1346 **arXiv:1412.6572 [cs, stat]**, 2014.
- 1347 52. Szegedy, C, Zaremba, W, Sutskever, I, Bruna, J, Erhan, D, Goodfellow, I, and Fergus, R, *Intriguing*
1348 *properties of neural networks*. **arXiv:1312.6199 [cs]**, 2013.
- 1349 53. Brendel, W and Bethge, M, *Approximating CNNs with Bag-of-local-Features models works*
1350 *surprisingly well on ImageNet*. **arXiv:1904.00760 [cs, stat]**, 2019.
- 1351 54. Geirhos, R, Rubisch, P, Michaelis, C, Bethge, M, Wichmann, FA, and Brendel, W, *ImageNet-*
1352 *trained CNNs are biased towards texture; increasing shape bias improves accuracy and*
1353 *robustness*. **arXiv:1811.12231 [cs, q-bio, stat]**, 2018.
- 1354 55. Vinken, K and de Beeck, HO, *Deep Neural Networks Point to Mid-level Complexity of Rodent*
1355 *Object Vision*. **bioRxiv**, 2020: p. 2020.02.08.940189.
- 1356 56. Leakey, R, *The origin of humankind*. 1996: Basic Books.
- 1357 57. Wekselblatt, JB, Flister, ED, Piscopo, DM, and Niell, CM, *Large-scale imaging of cortical dynamics*
1358 *during sensory perception and behavior*. **Journal of Neurophysiology**, 2016. 115(6): p. 2852-
1359 2866.
- 1360 58. Stirman, JN, Townsend, LB, and Smith, SL, *A touchscreen based global motion perception task for*
1361 *mice*. **Vision Research**, 2016. 127: p. 74-83.
- 1362 59. Kalatsky, VA and Stryker, MP, *New paradigm for optical imaging: temporally encoded maps of*
1363 *intrinsic signal*. **Neuron**, 2003. 38(4): p. 529-545.
- 1364 60. Yang, L, Lee, K, Villagracia, J, and Masmanidis, SC, *Open source silicon microprobes for high*
1365 *throughput neural recording*. **J Neural Eng**, 2020. 17(1): p. 016036.
- 1366 61. Hunter, JD, *Matplotlib: A 2D Graphics Environment*. **Computing in Science & Engineering**, 2007.
1367 9(3): p. 90-95.
- 1368 62. McKinney, W, *Data Structures for Statistical Computing in Python*. 2010: p. 6.
- 1369 63. Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer,
1370 P, Weiss, R, Dubourg, V, Vanderplas, J, Passos, A, Cournapeau, D, Brucher, M, Perrot, M, and
1371 Duchesnay, É, *Scikit-learn: Machine Learning in Python*. **Journal of Machine Learning Research**,
1372 2011. 12(Oct): p. 2825-2830.
- 1373 64. Van Der Walt, S, Colbert, SC, and Varoquaux, G, *The NumPy array: a structure for efficient*
1374 *numerical computation*. **Computing in Science & Engineering**, 2011. 13(2): p. 22-30.
- 1375 65. Michael, W, Olga, B, Drew, OK, Paul, H, Joel, O, Saulius, L, David, CG, Tom, A, Yaroslav, H, John,
1376 BC, Jordi, W, Julian de, R, Cameron, P, Stephan, H, Jake, V, Santi, V, Gero, K, Eric, Q, Pete, B,
1377 Marcel, M, Kyle, M, Alistair, M, Yoav, R, Thomas, B, Tal, Y, Mike Lee, W, Constantine, E, Clark, F,
1378 Brian, and Adel, Q, *mwaskom/seaborn: v0.9.0 (July 2018)*. 2018, Zenodo.
- 1379 66. Hunter, JN and Born, RT, *Stimulus-dependent modulation of suppressive influences in MT*. **J**
1380 **Neurosci**, 2011. 31(2): p. 678-86.
- 1381