Feedforward and feedback processing during figure-ground perception in mice

Ulf H. Schnabel[1*], Lisa Kirchberger[1*], Enny H. van Beest[1*], Sreedeep Mukherjee[1*], Areg Barsegyan[1], Jeannette A. M. Lorteije[2], Chris van der Togt[1], Matthew W. Self[1] & Pieter R. Roelfsema[1,3,4,#]

Affiliations:

[1]Department of Vision & Cognition, Netherlands Institute for Neuroscience, Meibergdreef 47, 1105 BA, Amsterdam, The Netherlands.

[2]Cognitive and Systems Neuroscience Group, Center for Neuroscience, Faculty of Science, Swammerdam Institute for Life Sciences, University of Amsterdam

[3]Department of Integrative Neurophysiology, Center for Neurogenomics and Cognitive Research, VU University, Amsterdam, The Netherlands.

[4]Department of Psychiatry, Academic Medical Center, Amsterdam, The Netherlands

* These authors contributed equally

Corresponding author: Pieter Roelfsema: p.roelfsema@nin.knaw.nl

Number of figures: 4

Conflict of Interest: The authors declare no competing financial interests.

One sentence summary: Figure-ground perception in mice depends on reciprocal interactions between the primary visual cortex and higher cortical areas, with specific roles for PV-, VIP- and SOM-positive interneurons.

Abstract

The segregation of figures from the background is an important first step in the analysis of a visual scene. Figure-ground segregation is thought to rely on interactions between the primary visual cortex (area V1) and higher visual areas. Upon presentation of a new image, the initial V1 responses reflect the information in the neurons' receptive fields (RFs), whereas later activity is context-dependent so that the representations of figures are enhanced relative to the background[1–3]. It is unknown if the figural response enhancement in V1 plays a role in perception and the mechanisms that produce it are not well understood. We trained mice to report a contrast-defined figure or a figure on a textured background that required figure-ground segregation, and optogenetically silenced V1 activity at different time-points. Suppression of early V1 activity interfered with both visual tasks, but suppression of the later activity selectively interfered with figure-ground perception. Using widefield imaging, we observed that figures also elicited stronger activity than the background in higher visual areas, and we used optogenetics to demonstrate that the extra activity is fed back to enhance the V1 figure representation. Within V1, figures increased the activity of pyramidal cells, parvalbumin (PV) and vasoactive intestinal peptide (VIP) positive interneurons but they decreased the activity of somatostatin-positive (SOM) interneurons. These results support the view that VIP interneurons inhibit SOM-interneurons to disinhibit the cortical column[4]. Our results demonstrate that figure-ground modulation of V1 neurons contributes to perception. We expect that the observed interactions between lower and higher brain regions generalize to other sensory modalities and that they will inform neurobiologically realistic models of scene perception.

Neurons at early processing levels of the visual system initially analyze the visual scene in a piecemeal manner. They carry out a local analysis of the image elements in their small receptive fields (RFs) and feed the visual information forward to higher visual areas. Neurons in higher areas have larger RFs and integrate information to represent more abstract features of the visual scene, including object category and identity[5,6]. Researchers hypothesized that the early feedforward processing phase is followed by a phase of recurrent interactions between lower and higher cortical areas that supports perceptual organization, a process that groups image elements of behaviorally relevant objects and segregates them from other objects and the background[7]. Figure 1a illustrates images that induce perceptual organization; a subset of the texture elements forms a figure which appears segregated from the background. Previous studies demonstrated that neuronal responses in area V1 that are elicited by figures are stronger than those elicited by background regions, even if the image elements in the receptive field are the same (compare the RF in stimulus 1 vs. 2 and 3 vs. 4 Fig. 1b)[1,8]. The influence of figure-ground perception on firing rates, which must originate from outside the RF, is known as figure-ground modulation (FGM). In many situations, FGM does not occur during the initial response elicited by the onset of a stimulus but after a delay, consistent with theories that FGM depends on feedback from higher visual areas or horizontal connections within V1[8,9]. In accordance with a role of feedback, FGM is stronger when a figure receives attention[10] or if it enters into the subject's awareness[11]. However, it is unknown whether FGM in V1 is just an epiphenomenon of processing in higher brain regions that feed back to V1 or if it is necessary for perception. Furthermore, elucidating the mechanisms causing FGM, which are not well understood, will give new insight into how the neural representations of simple features in lower brain areas and more abstract features in higher brain areas interact with each other during perception. Here we therefore studied figure-ground perception in mice and addressed the following questions: (1) Is FGM causally related to figure-ground perception? (2) What are the roles of the different interneurons types for FGM? (3) Does FGM also occur in higher cortical areas and is it fed back to cause FGM in V1?

We trained mice to report whether a circular grating (0.075 cycles/deg) or texture pattern was positioned on the left or right side of a display by licking one of two spouts (Fig. 1a,b), while their head was restrained. The figure was positioned on a grey background or on a similar pattern with the orthogonal orientation or a different phase so that the mice had to segregate the figure from a background. The average accuracy of the mice for the stimulus with a grey background was 89.7±6.6% (mean ± s.d.; N=10 mice) (Fig. 1c). It was 72.6±5.6% for a grating figure with an orientation orthogonal to the background (N=13), 70.8±2.1% if the figure was defined by a phase-difference (N=6) and 64.9±2.4% (N=3) for the texture patterns. These results imply that mice can

perform figure-ground segregation based on a variety of visual cues and can therefore be used to study this form of perception, in accordance with a previous study demonstrating that mice also readily generalize their performance to new types of figure-ground stimuli[12].

To examine V1 activity elicited by figure-ground stimuli, we recorded multi-unit activity (MUA) with multichannel silicon probes that spanned all cortical layers (N=75 recording sites in 3 mice, passively viewing the stimuli). For the figure-ground stimuli, we ensured that the stimulus in the RF was identical in the figure and ground conditions (Fig. 1b), so that activity differences could only be caused by contextual information, outside the RF. For comparison, we also presented the stimulus with a grey background so that the RF either fell on a grating or a homogeneously grey background. Activity elicited by figures was stronger than that elicited by backgrounds, even if the RF stimulus was constant (Fig. 1d). We quantified the FGM for individual recording sites with the d-prime, a measure of the reliability of the activity difference at individual electrodes on single trials (see Methods, time-window 100-500ms after stimulus onset). The average d-prime was 1.83 for the contrast stimulus and it was 0.56, 0.25 and 0.45 for the orientation-defined, phase-defined and texture stimuli, respectively (all ps < 0.001, Bonferroni corrected t-test). To determine FGM-latency we subtracted V1 activity elicited by the background from that elicited by the figure and fitted a curve to the response difference (lower panels in Fig. 1d; see Methods). The latency for the contrast stimulus was 60ms, which was shorter than that for the orientation-defined figure-ground stimuli (95ms, $p < 0.001$, bootstrap test). Latencies for the phase-defined (121ms) and texture figure-ground stimuli (116ms) were, in turn, longer than for the orientation-defined figure-ground stimulus (both ps≤0.001). We also measured FGM in V1 of four mice during performance of the figure detection task (Fig. S1). FGM on correct trials in the figure-detection task was similar to that during passive viewing and even on error trials, in which the mice licked the wrong spout.

To test if FGM is necessary for figure-ground perception, we expressed the inhibitory opsin GtACR2, a light activated chloride channel[13], in V1 pyramidal neurons in five mice. Blue light quickly and reliably inhibited the neuronal activity across all cortical layers (Fig. S2c,d) and we varied the onset of the blue light across trials so that we could examine the phases of the V1 response that are essential for reliable performance[14]. In the contrast detection task, silencing from 33 ms after stimulus onset onwards, which abolishes the entire V1 response, reduced the accuracy of the mice (Fig. 1f) ($p<0.05$ in each mouse; we ensured that the presence of blue light itself did not cause interference, Methods and Fig. S2a,b). In most animals, the accuracy remained above chance level, however, implying that a relatively low level of performance in the contrast detection task can be maintained by brain regions outside V1. The accuracy quickly recovered when we postponed optogenetic silencing until after the visually driven V1 response and it reached its half maximal value after 64ms (95%-

confidence interval 59-69ms) (Fig. 1f,g). In the orientation-defined figure-ground task, silencing of V1 reduced the accuracy to chance level and it only recovered to its half-maximal value when V1 silencing was postponed to 102ms (95%-confidence interval 96-111ms), while the required processing time was 136ms if the figure was defined by a phase-difference (95%-confidence interval 110-147ms; ps<0.05 for latency differences between contrast detection and figure-ground segregation, Fig. 1g). We conclude that the initial V1 response transient suffices for accurate performance in the contrast detection task, whereas later V1 activity is necessary for figure-ground segregation, implying that the FGM signal is read out by brain structures that select the appropriate motor response.

We used two-photon imaging to examine how the pyramidal cells and the different interneuron subclasses of the V1 cortical microcircuit contribute to FGM. Previous studies in monkeys have suggested that both the enhancement of activity elicited by figures[10,15] and the suppression of activity elicited by the background depend on feedback from higher visual cortical areas[16,17]. In mice, feedback connections from higher visual areas can exert a suppressive effect by targeting somatostatin-expressing inhibitory neurons (SOM)[4] (Fig. 2a) and a previous study demonstrated that SOM neurons indeed suppress V1 activity elicited by homogeneous image regions[18]. By contrast, feedback connections to V1 can also disinhibit the cortical column by targeting vasoactive intestinal peptide-positive (VIP) interneurons, which inhibit SOM cells[4,19–21] (Fig. 2a). We examined the calcium responses of V1 pyramidal cells (Fig. 2b) in awake transgenic mice expressing GCaMP6f in pyramidal cells. The mice passively viewed stimuli with figures defined by an orientation difference. In pyramidal cells, the figures elicited stronger calcium responses than did backgrounds, just as in the electrophysiological experiments (Fig. 1c,d) (N=227 cells in 7 mice, t(226)=2.1, p=0.03, time-window 500-1500ms). The time-course of the response was protracted due to the kinetics of the calcium sensor, precluding an accurate determination of the FGM latency. We imaged the activity of VIP-neurons, by expressing GCaMP6f in VIP-Cre mice and observed that figures elicited stronger VIP-neuron activity than backgrounds (N=153 neurons in 4 mice; paired t-test, t(152)=3.9, p<0.001) (Fig. 2e-g), consistent with the hypothesis that these neurons receive excitatory feedback from higher visual areas about salient visual stimuli[4]. To examine the activity of SOM cells, we expressed GCaMP6f in SOM-Cre mice and found a suppression of the figure response compared to the background response (N=157 neurons in 4 mice, t(156)=-2.1, p=0.04) (Fig. 2h-j). This result supports the hypothesis that the ground suppression is associated with extra activity of SOM-neurons, whereas the enhancement of the figure representation is associated with a release of SOM-inhibition[21] (Fig. 2a). We also examined the activity of parvalbumin positive (PV) cells, which are of the fast-spiking type and have been implicated in feedforward inhibition and control of the gain of

the cortical column[19,22]. We expressed GCaMP6f in V1 of PV-Cre mice and observed that figures evoked stronger responses than backgrounds (N= 159 neurons in 4 mice, $t(158)=12.7$, $p<0.001$) (Fig. 2k-m), in accordance with a previous study which demonstrated that fast-spiking interneurons exhibit relatively strong activity increases to salient stimuli[23].

To examine neuronal activity across the cortical areas that could provide feedback to V1[24,25], we used wide-field imaging in nine Thy1-GCaMP6f mice with a transparent skull[26]. We first measured the retinotopy to determine the boundaries between the areas and to align them with the Allen Brain Map (Fig. S3d, see Methods) and investigated FGM elicited by orientation (N=9) and phase-defined (N=5) figure-ground displays, during passive viewing. We used the measured retinotopy to map the V1 activity profile into visual field coordinates and we observed that the entire visual field region occupied by the figure was labeled with enhanced activity, again with a protracted time-course caused by the calcium-imaging method (time-window 150-300ms; $p<0.001$) (Fig. 3a,b; Fig. S3a,b). Orientation and phase-defined figures also elicited more activity than the background in areas LM, AL, RL, A, AM, PM and RSP (Fig. 3c) (ANOVA for d-primes across mice, followed by post-hoc t-tests). To examine if FGM increases during active perception, we trained five of these mice to perform the orientation figure-ground task during widefield imaging. FGM during correct figure-detection trials was stronger than during passive viewing in A and AM (two-way ANOVA with factors area and behavioral condition, interaction effect $F(9,36) = 2.3$, $p<0.05$; post-hoc tests, $ps<0.05$ for A and AM; Fig. 3d). We also compared the activity between correct and erroneous trials. Just as in our electrophysiological experiments, the figure representation was enhanced relative to the background when mice licked on the wrong side, but FGM was weaker during errors in V1, PM, AM, A and RL (two-way ANOVA with factors area and choice, interaction effect $F(9,36)=2.8$, $p<0.05$; post-hoc tests, $ps<0.05$; Fig. S3c). These results confirm that FGM is a robust phenomenon that also occurs during passive viewing. FGM is weaker during the errors in the task, which presumably impairs the readout of the figure location by brain areas involved in response selection.

So far, we demonstrated that the delayed V1 response is important for figure-ground perception and that FGM occurs in higher cortical areas, which are known to provide feedback to V1. However, FGM could also be generated locally, within V1. To examine whether feedback to V1 contributes to FGM, we optogenetically silenced higher visual cortical areas and examined the influence on V1 activity. We injected AAV-CaMKII-hGtACR2 medially (in areas AM and PM) and laterally from V1 (LM and AL), after identification of the V1 borders using population receptive field mapping, to ensure that the virus did not spread into V1 (Fig. 4a). We recorded the activity elicited by different figure-ground displays in V1, inhibiting neuronal activity in the higher areas in 50% of the trials by applying blue light, starting 200ms before stimulus onset. In the example neuron of Fig. 4b, silencing of higher

areas did not influence the initial response elicited by the onset of the stimulus but it decreased the FGM so that the d-prime was reduced from 0.59 to 0.12 (bootstrapped independent t-test, $p = 0.03$). We obtained similar results across the population of 76 recording sites in three mice (Fig. 4c). Silencing of the higher areas did not influence the initial V1 response (time-window 0-100ms; Fig. 4c), but the FGM decreased by 5%, 20% and 44% for the orientation defined, phase-defined and textured figure-ground stimuli, respectively (paired t-test, $p=0.008$, $p<0.001$ and $p<0.001$, respectively; Fig. 4d). Hence, a substantial fraction of the FGM in V1 is caused by feedback from the higher visual areas, where figures also elicit more activity than the background.

The present results provide new insights in how the interactions between V1 and higher visual cortical areas determine perception. Upon the appearance of a new image, the activity of V1 neurons exhibits a number of phases[8]. During an early feedforward phase, activity arrives from the retina and is propagated onwards to higher cortical areas. V1 activity during this phase sufficed for ceiling performance in the relatively simple contrast detection task, as if V1 acts as relay station that only needs to transmit information about the grating location to higher areas[27]. If the stimulus requires figure-ground segregation, the peak response is followed by a phase in which the V1 neurons exhibit FGM; the representation of all figural image elements is labeled with enhanced neuronal activity. In previous work, it remained unclear whether FGM is important for perception, but we now demonstrated that optogenetic V1 silencing during the late phase blocked figure-ground perception. This finding is in line with previous TMS studies in humans, although TMS produces weaker perceptual effects[28]. We identified a source for V1-FGM in higher visual areas LM, AL, RL, A, AM, PM and RSP, where neurons also exhibit FGM and project back to V1. These feedback influences caused a fraction of the V1-FGM, in particular for the phase-defined and texture-defined figure-ground stimuli. Accordingly, optogenetic silencing of higher areas suppressed FGM in V1[29], but hardly influenced the early V1 responses. Our results differ from previous studies that examined the early and late phases of neuronal responses in the barrel cortex of mice engaged in tactile detection tasks[30,31]. Suppression of the late activity in the barrel cortex interfered with the detection of a tactile stimulus whereas in the present study early V1 activity sufficed for contrast detection. This difference between results may be related to the use of relatively weak stimuli in the previous tactile detection task, which may have required amplification by the recurrent interactions between lower and higher areas of the somatosensory cortex.

Within V1, the enhanced activity of excitatory neurons was associated with a specific activation pattern of inhibitory interneurons. A previous study[4] suggested that feedback connections enhance V1 activity at relevant locations by activating of VIP interneurons, which in turn inhibit inhibitory SOM cells, thereby disinhibiting the cortical column. In line with this hypothesis, we observed that

figures enhanced the activity of VIP neurons and suppressed the activity of SOM cells. By contrast, the background increased the activity of SOM cells, which supports previous evidence that SOM interneurons suppress the V1 activity elicited by homogeneous image regions[18].

The present results provide insight into how interactions between lower and higher areas of the visual cortex enable the co-selection of image elements that belong to a single figure and their segregation from the background. We found that the representations of figural image elements are labeled with enhanced neuronal activity across multiple areas of the visual cortex, in accordance with previous studies suggesting that such a labeling process determines the perceptual grouping of features of the same object coded in different areas[32–35]. It is likely that our findings generalize to more complex forms of perceptual organization. To give an example, imagine grasping an object in a box filled with other objects. The visual system guides our fingers to touch and grasp edges of the same object, an attentional selection process[36] that has to rely on recurrent interactions between higher areas coding for the object's shape and lower areas coding for the individual edges. Previous studies in monkeys demonstrated that the representation of a selected object's shape is enhanced in higher cortical areas[37] and that this also holds true for relevant edges in V1[38,39]. The present study illustrates how this selection process can be coordinated across cortical areas. Future studies can therefore start to investigate the interactions between brain regions that are required for these more complex forms of perceptual organization and how they guide behavior[7].

Acknowledgments

We thank the animal caretakers at the Netherlands Institute of Neuroscience for their assistance. The work was supported by NWO (ALW grant 823-02-010) and the European Union's Horizon 2020 and FP7 Research and Innovation Program (grant agreement 7202070 ''Human Brain Project SGA1 and SGA2'', ERC grant agreement 339490 ''Cortic_al_gorithms'' and the Erasmus Mundus "NeuroTime" program) and the Stichting Vrienden van het Herseninstituut.

Author contributions

UHS, EvB and JAML developed the task. UHS, MWS and LK recorded and analyzed electrophysiological data. SM carried out the two-photon experiments and UHS, MWS, SM and CvdT analyzed the two-photon data. LK performed and analyzed optogenetics experiments with help from MWS. EvB and AB performed the wide-field experiments and EvB and MWS analyzed the results. PRR conceived of the experiments and supervised the project together with MWS. PRR, UHS, LK, SM and EvB wrote the paper.

Competing interests

The authors declare that they do not have competing interests.

Materials and correspondence

Correspondence and requests for materials can be sent to PRR.

References

1. Lamme, V. A. F. The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.* **15,** 1605–1615 (1995).

2. Self, M. W., van Kerkoerle, T., Supèr, H. & Roelfsema, P. R. Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Curr. Biol.* **23,** (2013).

3. Roelfsema, P. R. & de Lange, F. P. Early visual cortex as a multiscale cognitive blackboard. *Annu. Rev. Vis. Sci.* **2,** 131–151 (2016).

4. Zhang, S. *et al.* Long-range and local circuits for top-down modulation of visual cortex processing. *Science (80-.).* **345,** 660–665 (2014).

5. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18,** 193–222 (1995).

6. Yamins, D. L. K. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **19,** 356–365 (2016).

7. Roelfsema, P. R. Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* **29,** 203–227 (2006).

8. Lamme, V. A. F. & Roelfsema, P. R. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* **23,** (2000).

9. Klink, P. C., Dagnino, B., Gariel-Mathis, M. A. & Roelfsema, P. R. Distinct feedforward and feedback effects of microstimulation in visual cortex reveal neural mechanisms of texture segregation. *Neuron* **95,** 209–220 (2017).

10. Poort, J. *et al.* The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron* **75,** 143–156 (2012).

11. Supèr, H., Spekreijse, H. & Lamme, V. A. F. Two distinct modes of sensory processing observed in monkey primary visual cortex. *Nat. Neurosci.* **4,** 304–310 (2001).

12. Schnabel, U. *et al.* Figure-ground perception in the awake mouse and neuronal activity elicited by figure-ground stimuli in primary visual cortex. *Sci. Rep.*

13. Govorunova, E. G., Sineshchekov, O. A., Janz, R., Liu, X. & Spudich, J. L. Natural light-gated anion channels: a family of microbial rhodopsins for advance optogenetics. *Science (80-.).* **349,** 647–650 (2015).

14. Resulaj, A., Ruediger, S., Olsen, S. R. & Scanziani, M. First spikes in visual cortex enable perceptual discrimination. *Elife* **7,** 1–22 (2018).

15. Klink, P. C., Dagnino, B., Gariel-Mathis, M.-A. & Roelfsema, P. R. Distinct feedforward and feedback effects of microstimulation in visual cortex reveal neural mechanisms of texture-segregation. *Neuron* **95,** 209–220 (2017).

16. Schwabe, L., Obermayer, K., Angelucci, A. & Bressloff, P. C. The role of feedback in shaping the extra-classical receptive field of cortical neurons: a recurrent network model. *J. Neurosci.* **26,** 9117–9129 (2006).

17. Nassi, J. J., Lomber, S. G. & Born, R. T. Corticocortical feedback contributes to surround suppression in V1 of the alert macaque. *J. Neurosci.* **33,** 8504–8517 (2013).

18. Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J. & Scanziani, M. A neural circuit for spatial summation in visual cortex. *Nature* **490,** 226–231 (2012).

19. Harris, K. D. & Shepherd, G. M. G. The neocortical circuit: themes and variations. *Nat. Neurosci.* **18,** 170–181 (2015).

20. Pi, H.-J. *et al.* Cortical interneurons that specialize in disinhibitory control. *Nature* **503,** 521–524 (2013).

21. Karnani, M. M. *et al.* Opening holes in the blanket of inhibition: localized lateral disinhibition by VIP interneurons. *J. Neurosci.* **36,** 3471–3480 (2016).

22. Ji, X. -y. *et al.* Thalamocortical innervation pattern in mouse auditory and visual cortex: laminar and cell-type specificity. *Cereb. Cortex* **26,** 2612–2625 (2015).

23. Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* **55,** 131–141 (2007).

24. Osakada, F. *et al.* New rabies virus variants for monitoring and manipulating activity and gene expression in defined neural circuits. *Neuron* **71,** 617–631 (2011).

25. Wang, Q., Gao, E. & Burkhalter, A. Gateways of ventral and dorsal streams in mouse visual cortex. *J. Neurosci.* **31,** 1905–18 (2011).

26. Dana, H. *et al.* Thy1-GCaMP6 Transgenic Mice for Neuronal Population Imaging In Vivo. *PLoS One* **9,** e108697 (2014).

27. Thorpe, S., Fize, D. & Marlot, C. Speed of processing in the human visual system. *Nature* **381,** 520–522 (1996).

28. Wokke, M. E., Sligte, I. G., Steven Scholte, H. & Lamme, V. A. F. Two critical periods in early visual cortex during figure-ground segregation. *Brain Behav.* **2,** 763–777 (2012).

29. Hupé, J. M. *et al.* Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* **394,** 784–787 (1998).

30. Sachidhanandam, S., Sreenivasan, V., Kyriakatos, A., Kremer, Y. & Petersen, C. C. H. Membrane potential correlates of sensory perception in mouse barrel cortex. *Nat. Neurosci.* **16,** 1671–7 (2013).

31. Manita, S. *et al.* A top-down cortical circuit for accurate sensory perception. *Neuron* **86,** 1304–1316 (2015).

32. Duncan, J., Humphreys, G. K. & Ward, R. Competitive brain activity in visual attention. *Curr.Opin.Neurobiol.* **7,** 255–261 (1997).

33. Roelfsema, P. R. & Houtkamp, R. Incremental grouping of image elements in vision. *Attention, Percept. Perform.* **73,** 2542–2572 (2011).

34. Roelfsema, P. R. Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* **29,** 203–227 (2006).

35. O'Craven, K. M., Downing, P. E. & Kanwisher, N. fMRI evidence for objects as the units of attentional selection. *Nature* **401,** 584–587 (1999).

36. Baldauf, D. & Deubel, H. Attentional landscapes in reaching and grasping. *Vision Res.* **50,** 999–1013 (2010).

37. Chelazzi, L., Miller, E. K., Duncan, J. & Desimone, R. A neural basis for visual search in inferior temporal cortex. *Nature* **363,** 345–347 (1993).

38. Roelfsema, P. R. & Houtkamp, R. Incremental grouping of image elements in vision. *Attention, Perception, Psychophys.* **73,** (2011).

39. Van Kerkoerle, T., Self, M. W. & Roelfsema, P. R. Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat. Commun.* **8,** (2017).

40. Marshel, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional specialization of seven mouse visual cortical areas. *Neuron* **72,** 1040–54 (2011).

41. Guo, Z. V *et al.* Flow of cortical activity underlying a tactile decision in mice. *Neuron* **81,** 179–194 (2014).

42. Self, M. W. *et al.* Orientation-Tuned Surround Suppression in Mouse Visual Cortex. *J. Neurosci.* **34,** 9290–9304 (2014).

43. Pnevmatikakis, E. A. *et al.* Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. *Neuron* 285–299 (2015). doi:10.1016/j.neuron.2015.11.037

44. Pnevmatikakis, E. A. & Giovannucci, A. NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *J. Neurosci. Methods* **291,** 83–94 (2017).

Figure legends

Figure 1 | Figure-ground perception relies on late V1 activity. a, Stimuli that induce figure-ground perception. We presented stimuli for which the figure was a grating that differed from the background in orientation or phase, a texture stimulus, and a circular grating on a grey background that did not require figure-ground segregation but was included for comparison. **b,** The mice reported the location of a figure on the left or right by licking a spout on the corresponding side. Note that the image elements in the RF (green circle) were identical in the figure and background conditions. **c**, Average accuracy of the mice when performing the task with different stimuli. **d**, The left panel illustrates the silicon probe recordings in V1 and an example receptive field. The other panels show the MUA response elicited by a contrast-defined figure and the different figure-ground stimuli. The panels below show the difference in activity elicited by figure and ground and a curve that was fitted to the difference response to determine FGM latency. **e,** Optogenetic experiment in which laser light was directed to the center of V1 in both hemispheres. The mice could respond after 200ms and we silenced V1 activity at different delays after stimulus appearance. **f,** Accuracy of five mice in the task with different stimulus types, as function of the onset time of optogenetic V1 silencing. The data points on the right of the graphs are the accuracies on trials without silencing. The data of individual mice are show as grey/black symbols and curves are fits of logistic functions. Colored curves, fits to the average accuracy across mice. The inflection point of the curve was taken as measure of latency (the 95% confidence interval is shown above the abscissa). Asterisks above the plot indicate when the accuracy of the individual mice was above chance level ($p<0.05$, Bonferroni-corrected binomial test). **g**, V1 processing time necessary to reach half maximal performance for the different stimuli. Symbols show data from individual mice (same as in f) and the horizontal lines show the result based on the average accuracy. Error bars, s.e.m. determined by bootstrapping. The data points shown on the right are FGM latencies in V1 as determined by electrophysiology (panel d).

Figure 2 | The activity of interneurons and pyramidal cells in area V1 during figure-ground segregation. a, upper, Previous studies demonstrated that feedback connections from higher areas target SOM neurons to inhibit V1 pyramidal cells and VIP neurons that inhibit SOM cells to release SOM inhibition and thereby disinhibit the pyramidal neurons[4]. PV cells are thought to cause feedforward inhibition and control the gain of the cortical column[22]. **lower**, Our results reveal that figures enhance the activity of VIP-neurons, whereas background enhance the activity of SOM-

neurons. **b,** We imaged the activity of pyramidal neurons in in Thy1 mice expressing GCaMP6f. We visualized the responsive regions by computing correlation images; luminance values represent the average correlation between a pixel and its 8 neighbors, measured over 3,000 frames. Scale bar, 100μm. **c**, We determined the receptive field (red circle) and placed the figure (grey circle) or the background in the RF. The middle and right panels show the activity of an example pyramidal neuron and the average activity of the population of 227 pyramidal cells in V1 elicited by the figure-ground stimuli, after normalization. Average activity is shown as normalized units (n.u.). The shaded area shows ±S.E.M. **d,** The activity of the pyramidal cells after normalization, in a time window from 500-1500ms after stimulus onset. Figures elicited stronger responses than the background ($t(226)=2.1$, $p=0.03$). **e,** GCaMP6f expression in an example VIP-Cre mouse induced with AAV9-CAG-Flex-GCaMP6f. **f,** Activity of an example VIP neuron and a population of 153 VIP neurons in four mice elicited by the figure-ground stimuli. The lower panel illustrates the RF of the example cell. **g,** Figures elicited stronger responses in VIP neurons than the background ($t(152)=3.9$, $p<0.001$). **h,** The expression of GCaMP6f in an example SOM-Cre mouse after AAV injection in V1. We only imaged the activity of cell bodies (Methods). **i,** Calcium response of an example SOM cell and a population of 157 SOM neurons imaged in four mice. **j,** The calcium response elicited by the background was stronger, on average, than that elicited by the figure ($t(156)=-2.1$, $p=0.04$). **k,** Expression of GCaMP6f in V1 of an example PV-Cre mouse after AAV injection. We observed strong expression in the cell bodies. **l,** Activity of an example PV neuron and a population of 159 PV neurons in four mice elicited by the figure-ground stimuli. **m,** In PV neurons, figures elicited stronger calcium responses than the background ($t(158)=12.7$, $p<0.001$).

Figure 3 | Widefield imaging of activity across areas of the visual cortex. a, Difference in activity ($\Delta F/F$) for an orientation-defined figure presented on the right and left side of the screen (top panel) in a time window 150-300ms after stimulus onset during passive viewing in an example mouse (the data of the other mice are presented in Fig. S3a). Red (green) regions indicate significant higher calcium responses if the figure is on the right (left) ($p<0.05$; uncorrected). **b,** V1 activity elicited by the figure-ground stimuli. **left,** V1 activity elicited by an orientation defined figure-ground stimulus transformed in visual coordinates based on the retinotopy. The black circle illustrates the outline of the figure. **right,** Average time-course of the response elicited by figure for V1 pixels falling within figure representation (black curve - black outline in the left panel) and the background (grey curve). The shaded area denotes s.e.m. across 9 mice. **c, left,** We aligned the brains of individual mice to the Allen brain common coordinate framework (Methods). **right,** d-prime for orientation and phase defined figure-ground stimuli in passively viewing mice. *, $p<0.05$, **, $p<0.01$, ***, $p<0.001$. Error

bars denote s.e.m. **d,** D-prime for orientation defined figure-ground stimuli during passive viewing and hits in the figure-detection task. Error bars denote s.e.m.

Figure 4 | Contributions of higher visual areas to FGM in V1. a, Example field-sign map generated using population receptive field mapping overlaid on the visual cortex (Supplementary Methods). The black circle shows the location of an example electrode penetration, which was guided by the field-sign map and the blood vessel pattern. Viral vectors encoding for the inhibitory chloride channel GtACR2 were targeted to four points (triangles), at least 500µm from the boundary of V1 (dashed line). Right panel, expression profile in a dorsal view of the ex-vivo brain, imaged with a wide-field fluorescence microscope. **b,** MUA at an example V1 recording site elicited by the figure (purple curve) and background (black) of the texture stimulus without optogenetic silencing (left) and with the laser targeting higher visual areas (right panel). **c,** MUA during the early response phase (0-100ms) at 76 V1 units with (y-axis) and without optogenetic silencing of higher visual areas (x-axis) during passive viewing of orientation defined figures (left), phase defined figures (middle) and texture-defined figures (right). Insets in the right top corner show the distribution of differences between the two conditions. Red arrow, mean difference. **d,** D-prime values for the FGM with (y-axis) and without optogenetic silencing (x-axis) in a time-window from 150-500ms. Silencing the lateral visual areas reduced figure-ground modulation for all figure-ground stimuli.

Methods

Task

The mice that were trained in the figure-ground task were held on a reverse day-night cycle and a fluid restriction protocol with a minimal intake of 0.025ml/g, while their health was carefully monitored. The animals were trained to indicate the side on which a figure appeared at by licking the matching side of a custom made double lick spout (Fig. 1a,b). We registered licks by measuring a change in capacitance or a voltage drop using an Arduino. A trial started when the stimulus with a figure on the left or right side of the mouse appeared on the screen. The stimulus was displayed for 1.5 seconds. The exact figure location varied slightly depending on the experiment and the RFs of the neurons that were studied, but the figure center was generally close to an azimuth of ±30° (left or right of the mouse) and an elevation of 15°.

Stimulus presentation was followed by a variable inter-trial interval (ITI) of 6-8s. If the animal made an error, a 5s timeout was added to the ITI. We presented a background texture during the ITI but the mice did not receive a reward if they licked in the presence of this background texture. In some of the sessions we included correction trials, which were repeats of the same condition after an error. We only included the non-correction trials to compute accuracy, defined as hits/(hits+errors). During task performance in the widefield imaging experiments, we used a motor which moved the lick spout close to the mouth of the mouse, 500ms after the presentation of the stimulus, thereby enforcing a minimum viewing time before the mouse could respond. In the electrophysiological experiments, the mice had an average hit-rate of 70.0% in the orientation-defined figure-ground task. In the optogenetic V1 silencing experiments, the mice performed the contrast detection task with an accuracy of 90.2 ± 6.1% (s.d., 5 mice, only taking trials without the blue light into account), the orientation defined task with an accuracy of 70.4 ± 2.2% (4 mice) and the task in which the figure was defined by a phase difference with an accuracy of 69.4 ± 2.2% (3 mice). The accuracy of the mice that performed the orientation-defined figure-ground task in the widefield imaging experiment was 74.7% when we excluded long stretches of trials where the animals appeared to be distracted (see below). The accuracies of all individual mice can be found in Table S1.

To perform the figure-ground task, we first trained the mice to detect grating figures on a grey background. The luminance-defined background was then replaced by a grating with a gradually increasing contrast across a number of weeks of training. The experimental procedures complied with the National Institutes of Health Guide for Care and Use of Laboratory Animals and the protocol used was approved by the ethical committee of the Royal Netherlands Academy of Arts and Sciences and the CCD; all experiments were performed in accordance with the relevant guidelines and regulations.

Visual Stimuli

We created the visual stimuli with the Cogent toolbox of Matlab and corrected the luminance profile of the monitor for nonlinearity. For the behavioral electrophysiological experiments, the stimuli were presented on a 21-inch LCD screen with a resolution of 1280x720 pixels (Dell 059DJP) driven by a windows computer at 60 Hz at a distance of 15 cm in front of the mouse. During the two-photon experiments, we presented visual stimuli to the left eye of the mice, using a 24-inch LCD monitor (Dell U2414H) with a resolution of 1920x1080 pixels and a refresh rate of 60Hz, placed at an angle of 30 deg. relative to the nose and a distance of 12 cm from the eye. For the optogenetics experiments, we used a 24-inch LCD monitor (1920 x 1200 pixels, Dell U2412M), placed 11 cm in front of the eyes and for widefield experiments an LCD monitor (68 x 122cm, Iiyama LE5564S-B1), at a distance of 14

cm. We applied a previously described[40] correction for the larger distance between the screen and the mouse at higher eccentricities. This method defines stimuli on a sphere and calculates the projection onto a flat surface. The orientation defined and phase defined figure-ground stimuli were composed of 100% contrast sinusoidal gratings with a spatial frequency of 0.075 cycles/deg. and a mean luminance of 20cd/m^2. The size of the figure was either 35 deg. (optogenetic- and wide-field imaging experiments), 40 deg. (electrophysiology during passive viewing), 50 deg. (two-photon experiments and electrophysiology during task performance; we used 45 deg. in one of these mice). For the orientation defined figures (Fig. 1a), the grating orientation in the background was either horizontal or vertical and the orientation of the figure was orthogonal. For the phase defined figures (Fig. 1a), the phase of the figure grating was shifted by 180 deg. relative to that of the background. For the contrast-defined stimuli, we presented the figure gratings on a grey background (20cd/m^2). To test the generality of figure-ground perception, we also presented random textures in some of the experiments, again ensuring that the figure and background stimulus in the receptive field of the neurons was identical. The texture was made by filtering Gaussian distributed random noise patterns through an oriented filter. Four different random noise patterns were generated afresh for each experimental session, these were filtered with 0 and 90 deg. filters yielding eight different oriented textures. The oriented filter (F) was made through summation of individual Gabor filters (G) as follows:

$$F = \frac{\sum_{i=1}^{20} \frac{G_{f(i),\varphi}}{rms(G_{f(i),\varphi})}}{20} \qquad (1)$$

Where G is Gabor filter of spatial frequency *f(i),* where *f* was linearly spaced from 0.05 to 0.5 cycles/deg, orientation φ (0 or 90 deg.) and standard deviation *1/f*, *rms* indicates a root-mean square operation. After filtering, the texture was matched to the root-mean square contrast of the sine-wave grating used in the other experiments using an iterative hard-clipping procedure with 10 iterations.

Anesthesia during surgeries

The mice were handled five to ten minutes per day starting one week before surgery. Anesthesia was induced using 3-5% isoflurane in an induction box and it was maintained using 1.2-2.5% isoflurane in an oxygen enriched air (50% air, 50% O_2) mixture and we subcutaneously injected 5 mg/kg meloxicam (0.5 mg/ml) as general analgesic. The mice were positioned in a stereotactic frame and the depth of anesthesia was monitored by frequently checking paw reflexes and breathing rate. The temperature of the animal was monitored and kept between 36.5° and 37.5° with a heating pad coupled to a rectal thermometer. The eyes were covered with ointment to prevent dehydration. The

area of incision was shaved, cleaned with betadine and lidocaine spray was applied to the skin as a local analgesic.

Electrophysiology

The passive laminar electrophysiological recordings were carried out in six Tg(Thy1-GCaMP6f)GP5.17Dkim mice (2 male, 4 female) aged between two and six months. We implanted a bar to immobilize the animal's head in the experimental setup. After induction of general anesthesia (as described above), the skin on the skull was opened, tissue was removed from the skull and we briefly applied H_2O_2. We then applied a dental primer (Keerhawe Optibond) to improve the bonding of cement to the skull. We applied a thin layer of dental cement (Heraeus Charisma A1) across an area of the skull anterior to bregma and fixed a head bar with additional dental cement (Vivadent Tetric Evoflow). We implanted two small screws over the cerebellum, which served as reference and ground wires. We used dental cement (Vivadent Tetric Evoflow) to fix them in place and build a well around the skull covering the posterior part of the left hemisphere, to prevent growth of the skin over the area of interest. We applied a thin layer of adhesive to the bone which clarifies the skull, a method referred to as the 'clear skull cap' technique[41]. After one week of recovery, we used population receptive field mapping (see Supplementary Methods) to determine the borders of the visual areas. We used these borders to target virus injections to higher visual areas and the electrode penetrations to a V1 region with RFs at ~30 azimuth and 20 elevation. For virus injections in mice M38-M40 (see Table S2 for details), we anesthetized the animals as described above and made four small craniotomies, each at a minimum distance of 500 µm from the V1 border. We slowly lowered a pulled borosilicate capillary containing the virus (AAV5-CaMKII-hGtACR2-eYFP, titer $7.1*10^{12}$) to a depth of 600 µm from the brain surface and slowly injected a total of 40nL per injection site (15 nL at 600 µm, 10 nL at 400 µm, 15 nL at 200 µm) using a Nanoject III programmable injector (Drummond Scientific). After the injection, we left the pipette in place for five minutes before slowly retracting it to avoid efflux of virus. We sealed the chamber with the bio-compatible adhesive Kwik-Cast (WPI). Mice M35-M37 were injected with AAV5-CaMKII-Jaws for different experiments; data from these mice included here is without optogenetic intervention. After three weeks, we performed the next surgery in which we made a craniotomy centered on the area of V1 with a population receptive field at 30 deg. azimuth and 20 deg. elevation. In electrophysiological recordings during passive viewing, the mice were placed on a treadmill, head-fixed and allowed to run or sit freely. We inserted a linear-array recording electrode (A1x32-5mm-25-177, NeuroNexus, 32 channel probe, 25 micron spacing) in V1 and lowered it to around 1 mm below the brain surface and adjusted the depth of the electrode with reference to the current source density profile as

reported previously[42] to ensure coverage of all layers. We amplified the electrical signal from the electrodes and sampled it at 24.4 kHz using a Tucker-Davis-Technologies recording system. We removed muscle artifacts by re-referencing each channel to the average of all other channels before filtering the signal between 500 and 5000 Hz. We detected spikes by thresholding (positive and negative threshold) the band-passed signal at 4 times an estimate of the median absolute deviation and convolved the detected spikes with a Gaussian with a standard deviation of 1.3ms (and an integral of 1) to derive an estimate of multi-unit spike-rate. First, we measured the receptive field of the units recorded at each electrode. We presented the mouse with a sparse noise stimulus consisting of 4 white checks (8 by 8 deg., 40 cd/m^2) on a black background presented for 250ms with a 250ms inter-trial interval. The checks (>30 presentations per check) were positioned on a grid ranging from -64 to 16 deg. horizontally and -22 to 66 deg. vertically relative to the mouse's nose with negative values indicating the right hemifield and corrected for flat screen distortion as was described above. We averaged the MUA response evoked by each check in a time window from 50-400 ms after stimulus onset to obtain a map of visual responsiveness and fit a 2D-Gaussian to estimate the width and center of the receptive field (Fig. 1d; Fig. S1c). The quality of the fit was assessed using r^2 and a bootstrapped variability index (BVI), which estimated the reliability of the estimate of the RF center. We resampled an equal number of trials as in the original dataset (with replacement) and regenerated the Gaussian fit. The BVI is the ratio of the standard deviation of the RF center position and the standard deviation of the fitted Gaussian. Recording sites with a BVI larger than 1 were removed from the analysis. To measure FGM, we positioned a 40 deg. diameter figure so that it was centered on the RFs of layer 4 units in the right hemifield. To create the background condition, we shifted the figure to the left by 50-60 deg. so that the RF fell on the background. We only included recording sites if (1) the receptive field center was located within 15 deg. of the center of the figure, (2) the RF was reliable (r^2 of the Gaussian fit > 0.3, BVI < 1) and (3) if the signal-to-noise ratio was greater than 1 (SNR was the ratio of the mean activity between 0-100ms after stimulus onset to the standard deviation of the baseline activity (-200-0ms) across trials). These criteria led to the inclusion of 75 recording sites used for the data of Figure 1d,g (53.2% of contacts in mice M35-37) and 76 recording sites used for the data of Figure 4 (55.7% of contacts in mice M38-40). The figure-ground stimuli of the different types (orientation-, phase- or texture-defined) were presented in blocks of 32 stimuli. The position of the figure and the orientation of the underlying pattern/texture was pseudorandomized within the block. FGM was quantified with the d-prime, which is a measure for the reliability of the signal on individual trials (time window 100-500ms):

$$d_prime = \frac{\mu_F - \mu_G}{\sqrt{\frac{1}{2}(\sigma_F^2 + \sigma_G^2)}} \quad (2)$$

Where μ_F and μ_G are the means and σ_F^2 and σ_G^2 are the variances of the figure and ground response, respectively. As in previous studies, we estimate the latency of figure-ground modulation by fitting a function (described fully in ref. 10) to the difference in activity elicited by the figure and the background (time-window, from 0 to 300ms after stimulus onset). Briefly, the function is the sum of an exponentially-modulated Gaussian and a cumulative Gaussian, capturing the Gaussian onset of neural modulation across trials/neurons and the dissipation of modulation over time. The latency was then taken as the (arbitrary) point at which the fitted function reached 33% of its maximum value. Statistical tests on the latency estimates were performed by bootstrapping. We selected a random set of recording sites (with replacement) 1000 times and recalculated the latency estimate for each resampled population. In the experiments in which we optogenetically silenced the higher areas while recording from V1 (Fig. 4), we randomly interleaved 50% of trials with the laser switched on. We used a DPSS Laser (BL473T3-100FC, Shanghai Lasers & Optics Century Co.) emitting blue light (wavelength 473nm) with a constant power of 5 mW as a light source and directed the light at lateral higher visual areas through an optical fiber with a thickness of 200 µm. The laser was turned on 200 ms before onset of the visual stimulus, it remained on during the stimulus presentation time of 500 ms and was turned off 100 ms after that.

Optogenetic silencing of V1 during contrast and figure-ground perception

For optogenetic silencing of V1 during behavior we injected five male and one female C57BL/6 mice aged between 2 and 14 months with a cell-specific viral vector (AAV5-CaMKII-hGtACR2-eYFP) coding for the inhibitory GtACR2 opsin[13]. The general surgical procedures were described above. For the viral injections, an incision in the skin was made and the skin was gently pulled aside, exposing the area of the skull above the cortex. We used a dental drill to make small craniotomies above primary visual cortex of both hemispheres 2.7mm lateral from the midline, 0.5 mm anterior of lambda. We placed a pulled borosilicate capillary containing the virus vertically above the craniotomy touching the dura. Slowly the pipette was lowered to a depth of 600 µm from the brain surface and a total of 150nL per hemisphere with a concentration of $3*10^{11}$ GC/mL was slowly injected, at different depths using Nanoject III programmable injector (Drummond Scientific). After the injection, we left the pipette in place for at least 8 minutes before slowly retracting it to avoid efflux of virus. After two weeks, we carried out another surgery, in which we attached a head-fixation bar to the skull, as was described above (electrophysiological procedures). We applied a thin layer of adhesive to the bone which clarifies the skull, a method referred to as the 'clear skull cap' technique[41]. After the adhesive

dried, we applied a thin layer of transparent dental cement (C&B Super-Bond) on top. We added a small rim of dental cement to the outer edges of the skull cap to prevent growth of the skin over the area of interest. Before we started the optogenetic silencing experiments, we shortly anesthetized the mice with isoflurane (see above) to apply a thin layer of transparent nail polish (Electron Microscopy Sciences) to the cap to reduce light glare.

In the combined behavioral and optogenetic experiments, the mice could respond immediately by licking after the stimulus appeared on the screen. Once the animals performed the task consistently with an accuracy larger than 65%, we introduced trials in which neural activity in the visual cortex was inhibited by activating the opsin. We presented figure-ground stimuli (defined by orientation or phase) on 75% of all trials and, optogenetic silencing was applied in a random 25% of those trials. We presented contrast defined stimuli in 25% of all trials and because the accuracy was high and stable, we increased the fraction of trials with optogenetic intervention up to 50% (we also increased the ratio of trials in which the laser was turned on during control experiments when the laser was not directed at V1; Fig. S2a,b). We used a DPSS Laser (BL473T3-100FC, Shanghai Lasers & Optics Century Co.) emitting blue light (wavelength 473nm) as a light source and directed the light to the cortex with a split optical fiber (2x2 Coupler, 50:50, 200 µm, FC/PC to 1.25mm Ceramic Ferrule, Thorlabs, Newton). The two fiber ends were directed at the centers of area V1 in the two hemispheres. Optogenetic stimulation lasted for 2s with a constant power of 10 mW at each fiber end. The onset of stimulation was shifted relative to the presentation of the visual stimulus in steps of 16.7ms according to the frame rate of the screen. The time between stimulus onset and laser onset was 17, 33, 50, 67, 83, 100, 150 or 200ms. We took care to remove spurious cues that might be caused by the optogenetic stimulation. We placed a light shield around the head of the mouse to prevent light from the optogenetic stimulation reaching the eye and we presented blue light flashes at random times by driving a blue LED to flash every 0.5-1s with an Arduino in all optogenetic experiments. We only included sessions in which the overall accuracy of the mice was above 75% or the accuracy in both figure-ground and contrast-detection trials was above 65%. We used correction trials, which were repeats of the same condition after an error, to decrease a response bias and excluded these trials when computing accuracy (correction trials were not used after optogenetic silencing trials). We used binomial tests (with Bonferroni-correction) to examine whether performance was significantly above chance level during trials with optogenetic manipulation and also to test whether optogenetic inhibition impairs performance. To estimate the timing at which the accuracy reaches its half maximum, we fit a logistic function to the accuracy as a function of the laser onset latency using the Palamedes toolbox in Matlab. We used bootstrapping (1000 times) by sampling trials with replacement to determine the 95%-confidence interval of the latency.

Two-photon imaging

We carried out experiments with four VIP-Cre, four SOM-Cre and four PV-Cre animals to image the activity of interneurons and seven Thy1-5.17-GCaMP6f mice to image the activity of excitatory V1 neurons. The mice were between 2-6 months old, weighed between 20 and 35g and included both sexes (see Table S2 for details). The animals were anesthetized as described above. An incision in the skin was made along the anteroposterior axis, and the skin was gently pulled to the side, exposing the area of the skull above the cortex and the area posterior to lambda. For labelling V1, a small craniotomy was performed over center of right V1 (AP 0.5mm anterior to lambda and 2.5mm lateral from the midline). We slowly injected 200-300nL of AAV9-CAG-Flex-GCaMP6f at 20nl/min distributed across three depths (500µm, 300µm and 150µm from the pial surface), as described above. The craniotomy was sealed and the skin was sutured. Two weeks later, the animals underwent a second surgery to implant a head-ring for immobilization and a cranial window to allow imaging of neuronal activity. An incision in the skin was made in anteroposterior axis, and the skin was gently pulled to the side, exposing the area of the skull above the cortex and the area posterior to lambda. The bone of the target area was cleaned by removing remaining tissue with blunt dissection and briefly applying H_2O_2 after which dental primer was applied (Keerhawe Optibond). A circular metal ring was fixed on the skull with light cured dental cement, centered on the visual cortex of the right hemisphere parallel to the imaging plane. We made a circular craniotomy with a diameter of 4-5mm centered at 0.5mm anterior to lambda and 2.5 lateral from the midline. We carefully thinned the bone along the outer diameter of the craniotomy and slowly lifted the bone flap without damaging the dura, which was kept moist with warm ACSF or saline. The craniotomy was closed with a double layered glass coverslip, with the outer glass resting on the skull, which was fixed with dental cement (Vivadent Tetric Evoflow). After two weeks, we started to habituate the mice to head immobilization while they could run on a running belt under the two-photon microscope (Neurolabware) with a 15.7Hz frame rate. We imaged frames (512 x 764 pixels) through a 16X water immersion objective (Nikon, NA 0.80) at 1.7X zoom at a depth of 120-300µm. We targeted V1 based on the retinotopic maps obtained with one-photon retinotopic mapping (see below). A Ti-Sapphire pulsed laser (MaiTai, Spectra Physics) was tuned to 920 nm for delivering excitation light. First, we mapped the receptive field locations of the neurons. We presented 12 x 12 deg. white (38 cd/m^2) squares on a black background (0.05 cd/m^2) in an area ranging from -18 to 78 deg. horizontally and -21 to 51 deg. vertically relative to the mouse's nose. Each square was flashed twice for 166 ms with a blank interval of 166 ms, which was followed by delay of 500ms. The square at each location was presented 20 times. We calculated receptive fields based on the response in a

500ms window after stimulus onset. We fit a linear regression model to estimate the responses to the squares of the grid, regressing out the influence of running and the interaction between the visual stimulus and running. We fit a circular 2D-Gaussian to the beta values for every location to estimate the receptive field center and its full width at half maximum response strength. We evaluated the quality of the fit using the r^2 value and the BVI (see above; r^2 of the Gaussian fit > 0.3, BVI < 1, and a positive visual response). For the figure-ground stimuli, we presented 50 deg. figures defined by gratings of 6 different orientations with a maximum luminance of 38 cd/m^2 and a minimum luminance of 0.05 cd/m^2. The figure was presented 120 times each in two positions on the screen (20 repetitions per orientation), one centered on the receptive field location of a cluster of imaged cells, and at another position at a distance of 55 deg. The stimuli were presented in randomized order in blocks of 48 trials for 0.5 s with an ITI of 2.5 s to allow enough time the calcium signal of the previous trial to decay. We ensured that the stimulus in the RF was identical in the figure and background conditions, as described in the main text. We used CAIMAN[43,44] for pre-processing of the recorded signal. We performed rigid motion correction for small shifts in the data due to motion of the animal, followed by the extraction of ROIs and the $\Delta F/F$. We inspected fitted ROIs using custom written software to ensure that we only selected cell bodies. To generate population responses, we averaged across the responses of the cells. We assessed the significance of differences between responses elicited by the figure and ground for the different neuronal types with paired t-tests.

Wide-field imaging

We imaged 8 male and 1 female Thy1-5.17 GCaMP6f mice[26] aged between 2 and 14 months. The skull of the animals was prepared with the clear skull cap technique (described above). We placed the mice under a wide-field fluorescence microscope (Axio Zoom.V16 Zeiss/Ceanotech) to image a large part of the cortical surface. Images were captured at 20 Hz by a high-speed sCMOS camera (pco.edge 5.5) and recorded using the Encephaelus software package (Ceanotech). The visual stimuli were presented on a LCD monitor (68 x 122cm, Iiyama LE5564S-B1), placed 14 cm in front of the mouse. We monitored the size and position of the right pupil (50Hz sampling rate) and movements of the mouse with a piezo plate under the front paws of the mouse (50Hz sampling rate) to remove trials with large body or eye movements. We captured images of 1600x1600 pixels (~15µm per pixel) and down-sampled them to 400x400 pixels and applied a Gaussian mean filter smoothing kernel of 5x5 pixels. We carried out a population receptive field mapping session to determine the borders of primary visual cortex (V1) and the higher visual areas and we matched these areas to the Allen Brain common coordinate framework (Fig. S3d). We computed the average $\Delta F/F$ (relative to the baseline

fluorescence in a 300ms window before stimulus onset) of all pixels within an area. Trials were removed if (1) the mouse had a performance below 65% in a window of fifteen trials, (2) the mouse movements measured by the piezo element were stronger than a certain threshold (time-window from the start of a trial until 300ms after stimulus onset), or (3) the absolute z-score of $\Delta F/F$ was larger than 3.5 (removal of trials with artifacts). We measured FGM with the d-prime in a time window of 150-300ms after stimulus onset, averaged across both hemispheres. We assessed significance with a repeated measures ANOVAs, testing whether the d-prime (the intercept) was higher than 0. In addition, we used a repeated measures ANOVA to test the effect of area (i.e. V1, LM, AL, RL, A, AM, PM, RSP, M1 and M2), response type (hit, error or passive viewing) and figure condition (orientation defined or phase defined). Imaging during passive viewing was done before the training of the mice (N=5). Information about the mice that contributed to the different experiments can be found in Tables S1 and S2.

Data availability

All data and the computer code used to analyze the data will be made available for download and curated at the Human Brain Project Joint Platform.

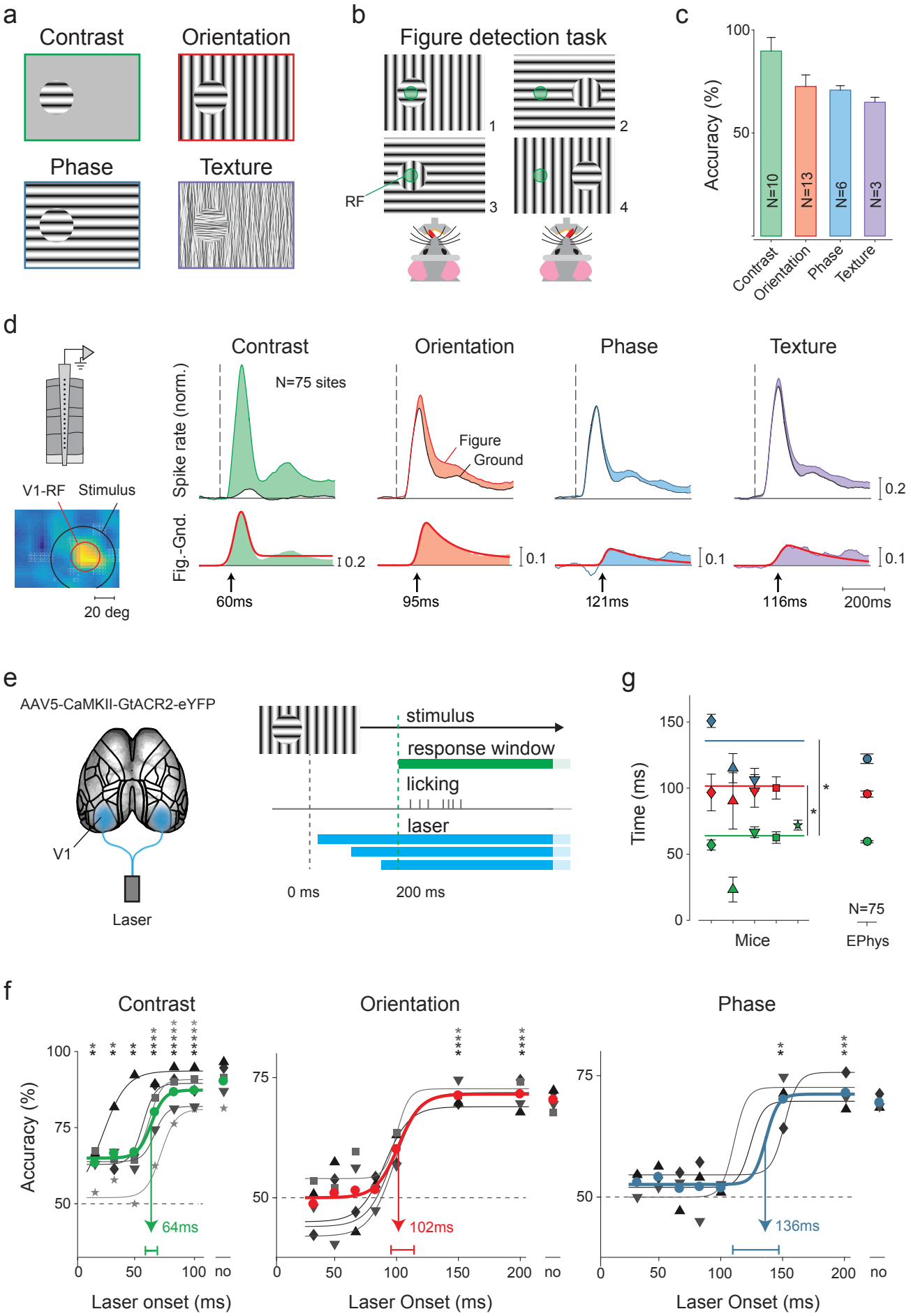
a
Contrast
Orientation
Phase
Texture
b
Figure detection task
1
2
3
4
RF
c
Accuracy (%)
100
50
N=10
N=13
N=6
N=3
Contrast
Orientation
Phase
Texture
d
Contrast
Orientation
Phase
Texture
Spike rate (norm.)
N=75 sites
Figure
Ground
0.2
Fig.-Gnd.
0.2
0.1
0.1
0.1
60ms
95ms
121ms
116ms
200ms
V1-RF
Stimulus
20 deg
e
AAV5-CaMKII-GtACR2-eYFP
V1
Laser
stimulus
response window
licking
laser
0 ms
200 ms
g
Time (ms)
150
100
50
0
*
*
Mice
N=75
EPhys
f
Contrast
Orientation
Phase
Accuracy (%)
100
75
50
64ms
102ms
136ms
0
50
100
no
150
200
Laser onset (ms)
Laser Onset (ms)

Figure 1

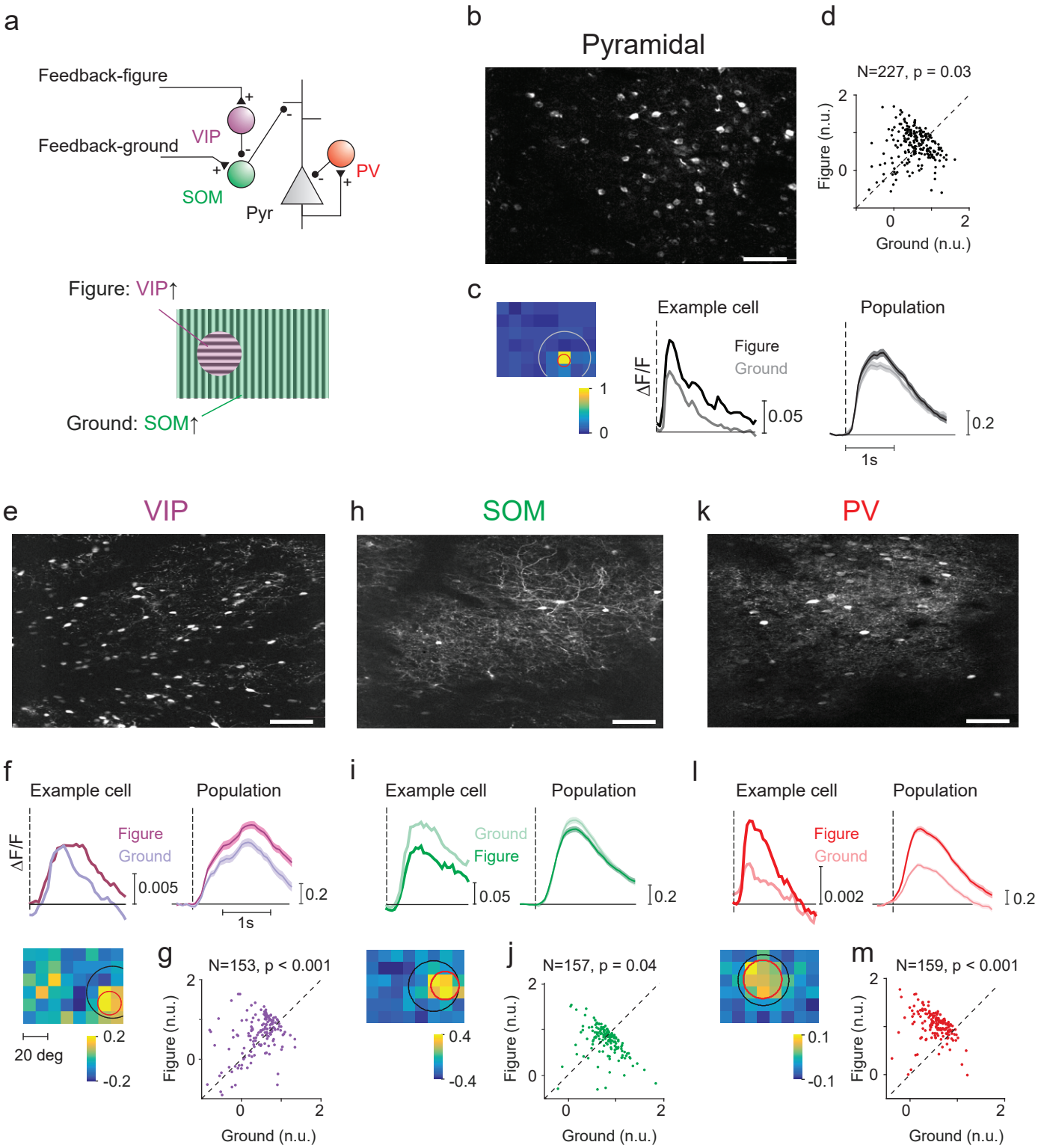
a
Feedback-figure
Feedback-ground
VIP
SOM
Pyr
PV
Figure: VIP↑
Ground: SOM↑
b
Pyramidal
d
N=227, p = 0.03
Figure (n.u.)
Ground (n.u.)
c
Example cell
Population
ΔF/F
Figure
Ground
0.05
0.2
1s
e
VIP
h
SOM
k
PV
f
Example cell
Population
ΔF/F
Figure
Ground
0.005
0.2
1s
20 deg
0.2
-0.2
g
N=153, p < 0.001
Figure (n.u.)
Ground (n.u.)
i
Example cell
Population
Ground
Figure
0.05
0.2
0.4
-0.4
j
N=157, p = 0.04
Figure (n.u.)
Ground (n.u.)
l
Example cell
Population
Figure
Ground
0.002
0.2
0.1
-0.1
m
N=159, p < 0.001
Figure (n.u.)
Ground (n.u.)

Figure 2

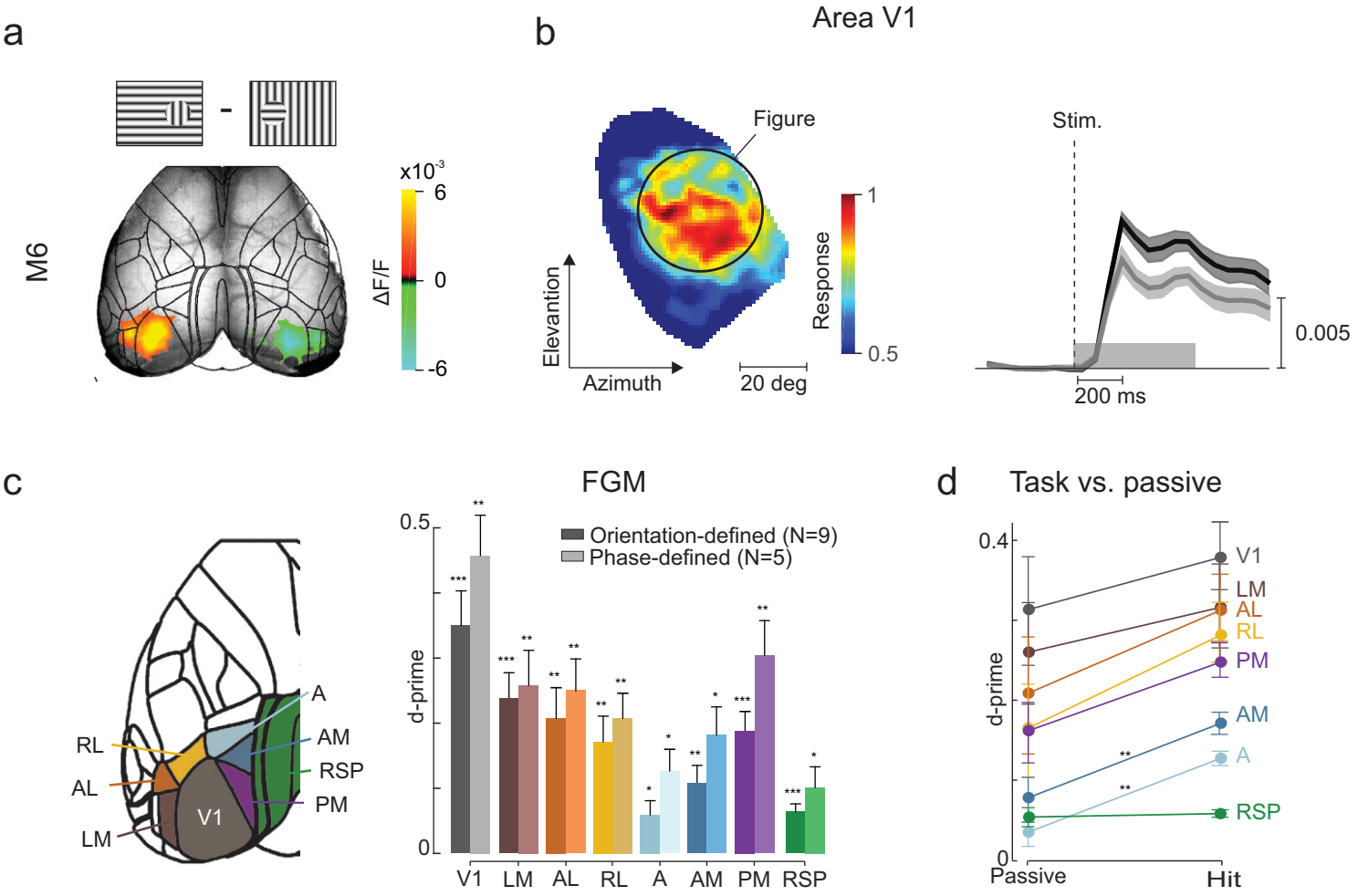
a
M6
x10-3
6
0
-6
ΔF/F
b
Area V1
Figure
Elevantion
Azimuth
20 deg
Response
1
0.5
Stim.
0.005
200 ms
c
FGM
Orientation-defined (N=9)
Phase-defined (N=5)
d-prime
0.5
0
V1
LM
AL
RL
A
AM
PM
RSP
d
Task vs. passive
0.4
0
Passive
Hit

Figure 3

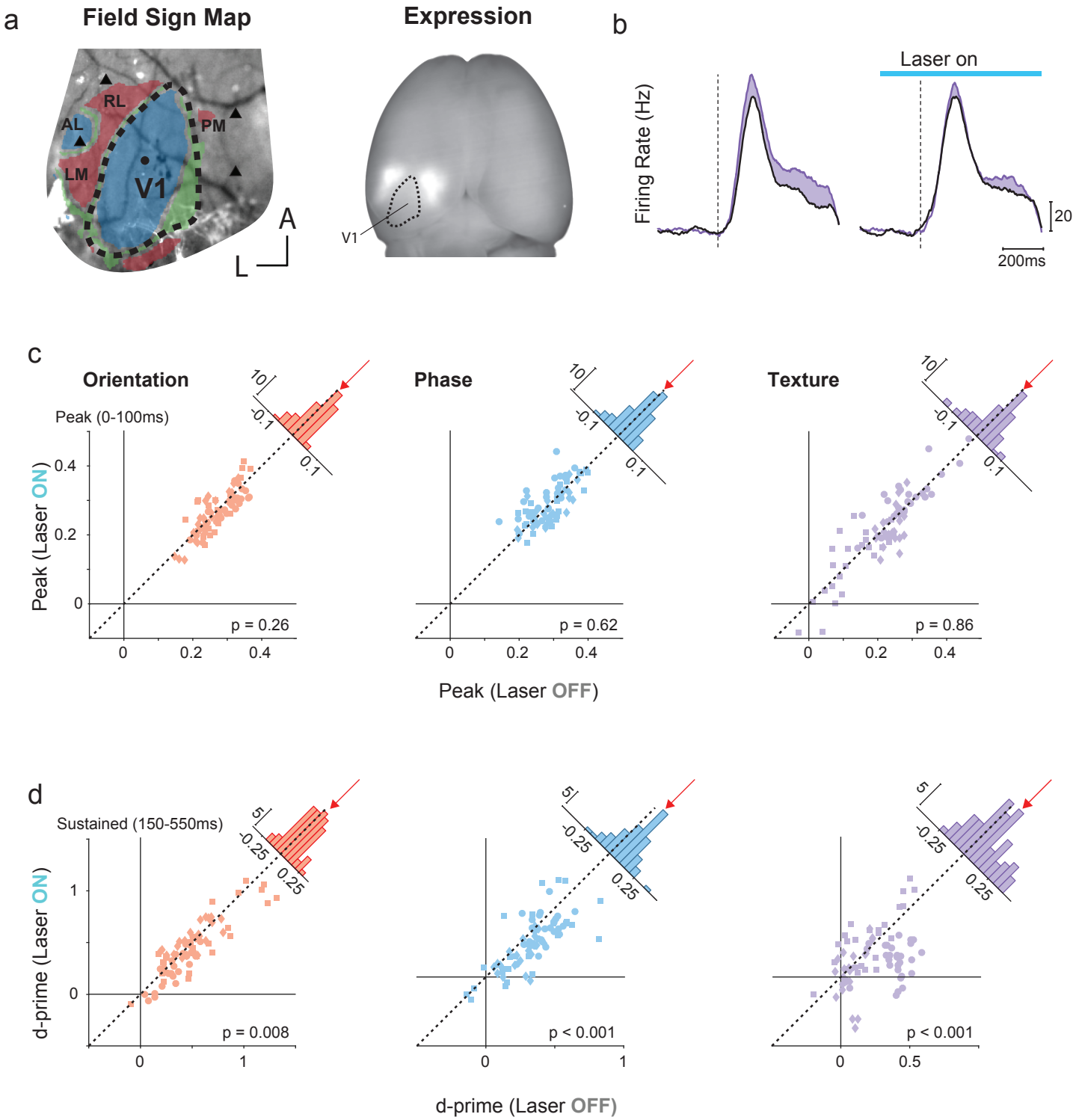
a
Field Sign Map
Expression
RL
AL
PM
LM
V1
A
L
V1
b
Laser on
Firing Rate (Hz)
20
200ms
c
Orientation
Phase
Texture
Peak (0-100ms)
Peak (Laser ON)
Peak (Laser OFF)
p = 0.26
p = 0.62
p = 0.86
d
Sustained (150-550ms)
d-prime (Laser ON)
d-prime (Laser OFF)
p = 0.008
p < 0.001
p < 0.001

Figure 4