

Learning from unexpected events in the neocortical microcircuit

Colleen J. Gillon^{1,2,3 †}, Jason E. Pina^{4,5 †}, Jérôme A. Lecoq^{6 †}, Ruweida Ahmed⁶, Yazan N. Billeh⁶, Shiella Caldejon⁶, Peter Groblewski⁶, Timothy M. Henley^{4,5}, India Kato⁶, Eric Lee⁶, Jennifer Luviano⁶, Kyla Mace⁶, Chelsea Nayan⁶, Thuyanh V. Nguyen⁶, Kat North⁶, Jed Perkins⁶, Sam Seid⁶, Matthew T. Valley⁶, Ali Williford⁶, Yoshua Bengio^{3,7,8}, Timothy P. Lillicrap^{9,10}, Blake A. Richards^{3,8,11,12 *}, and Joel Zylberberg^{4,5,8,13 *}

¹Department of Biological Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada

²Department of Cell & Systems Biology, University of Toronto, Toronto, Ontario, Canada

³Mila, Montréal, Québec, Canada

⁴Department of Physics and Astronomy, York University, Toronto, Ontario, Canada

⁵Centre for Vision Research, York University, Toronto, Ontario, Canada

⁶Allen Institute for Brain Science, Seattle, WA, USA

⁷Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Québec, Canada

⁸Learning in Machines and Brains Program, Canadian Institute for Advanced Research, Toronto, Ontario, Canada

⁹DeepMind, Inc., London, UK

¹⁰Centre for Computation, Mathematics and Physics in the Life Sciences and Experimental Biology, University College London, London, UK

¹¹School of Computer Science, McGill University, Montréal, Québec, Canada

¹²Department of Neurology & Neurosurgery, McGill University, Montréal, Québec, Canada

¹³Vector Institute for Artificial Intelligence, Toronto, Ontario, Canada

^{†,*}These authors contributed equally.

*Corresponding authors: blake.richards@mila.quebec (BAR), joelzy@yorku.ca (JZ)

Nov, 2021

Abstract

Scientists have long conjectured that the neocortex learns the structure of the environment in a predictive, hierarchical manner. According to this conjecture, expected, predictable features are differentiated from unexpected ones by comparing bottom-up and top-down streams of information. It is theorized that the neocortex then changes the representation of incoming stimuli, guided by differences in the responses to expected and unexpected events. In line with this conjecture, different responses to expected and unexpected sensory features have been observed in spiking and somatic calcium events. However, it remains unknown whether these unexpected event signals occur in the distal apical dendrites where many top-down signals are received, and whether these signals govern subsequent changes in the brain's stimulus representations. Here, we show that both somata and distal apical dendrites of cortical pyramidal neurons exhibit distinct unexpected event signals that systematically change over days. These findings were obtained by tracking the responses of individual somata and dendritic branches of layer 2/3 and layer 5 pyramidal neurons over multiple days in primary visual cortex of awake, behaving mice using two-photon calcium imaging. Many neurons in both layers 2/3 and 5 showed large differences between their responses to expected and unexpected events. Interestingly, these responses evolved in opposite directions in the somata and distal apical dendrites. These differences between the somata and distal apical dendrites may be important for hierarchical computation, given that these two compartments tend to receive bottom-up and top-down information, respectively.

Keywords: Learning; Neocortex; Pyramidal Neurons; Distal Apical Dendrites; Prediction; Credit Assignment; Deep Learning; Unsupervised Learning

45 1 Introduction

46 A long-standing hypothesis in computational and systems neuroscience is that the neocortex learns a hierarchical
47 predictive model of the world [Dayan et al., 1995; Friston and Kiebel, 2009; Hawkins and Blakeslee, 2004;
48 Larochelle and Hinton, 2010; Press et al., 2020; Rao and Ballard, 1999; Spratling, 2017; Whittington and Bogacz,
49 2017]. This hypothesis postulates that learned top-down predictions (i.e., signals from associative regions to
50 sensory regions) are compared to bottom-up signals (i.e., signals from sensory regions to associative regions)
51 (Fig. 1A). Unexpected stimulus events should then induce differences between these signals and, in turn, drive
52 learning. In these models, learning occurs at all stages of the hierarchy, and not just at the earliest or latest
53 stages. Theoretical support for this hypothesis comes from computational studies showing that hierarchical
54 models that learn by comparing top-down signals to bottom-up signals enable artificial neural networks (ANNs)
55 to learn useful representations that capture the statistical structure of the data on which they are trained [Chen
56 et al., 2020; Devlin et al., 2018; Grill et al., 2020; Lotter et al., 2016; van den Oord et al., 2018; Wayne et al.,
57 2018]. Moreover, ANNs trained in this manner reproduce the representations observed in the neocortex better
58 than ANNs trained purely by supervised learning based on categorical labels [Bakhtiari et al., 2021; Christensen
59 and Zylberberg, 2020; Higgins et al., 2017].

60 What would be the observable signatures of this type of hierarchical predictive learning, in which learning
61 is guided by unexpected sensory events? There are at least three signatures that one might expect: (1) There
62 should be distinct responses to expected and unexpected stimuli. If the brain does not distinguish between
63 expected and unexpected events, there is no way to specifically learn from the *unexpected* events. (2) As the
64 circuit learns about stimuli, the responses to both expected and unexpected stimuli should change in a long-
65 lasting manner. These changes in stimulus responses are a necessary consequence of learning modifying the
66 stimulus representations. (3) There should be differences between the manner in which top-down and bottom-
67 up driven responses change during learning. This follows from the idea that a hierarchical model is being
68 learned, since hierarchy implies a distinct role for top-down and bottom-up information.

69 Previous work has provided partial evidence for these observable signatures of hierarchical predictive learning.
70 First, there is a very large body of work showing distinct responses to expected and unexpected stimuli in
71 multiple species and brain regions [Fiser et al., 2016; Garrido et al., 2009; Keller et al., 2012; Kumaran and
72 Maguire, 2006; Orlova et al., 2020; Zmarz and Keller, 2016], thus supporting the first observable signature.
73 However, there are still significant unknowns: e.g., do such responses evolve differently in different compartments
74 of neurons? Second, there is some research suggesting that responses to unexpected stimuli change with exposure
75 [Homann et al., 2017], supporting the second observable signature. Yet, this has only been shown over short time
76 scales, such as a single experimental session. Third, there are a few studies showing that top-down projections
77 carry distinct information to sensory areas [Fiser et al., 2016; Jordan and Keller, 2020; Orlova et al., 2020],
78 partially supporting the third observable signature. Nonetheless, it remains unknown whether changes in neural
79 responses driven by top-down versus bottom-up signals show distinct changes over learning. Thus, the goal of
80 this paper is to fill these gaps by concretely looking for all three of these signatures together in a systematic
81 study.

82 Here, we performed chronic two-photon calcium imaging of layer 2/3 and layer 5 pyramidal neurons at
83 both the cell bodies and the distal apical dendrites in the primary visual cortex of awake, behaving mice over
84 multiple days (Fig. 1B–D; Supp. Videos 1–3). These imaging planes were chosen since top-down signals largely
85 impinge on the distal apical dendrites within cortical layer 1, while bottom-up signals largely impinge on the
86 perisomatic compartments in deeper layers [Budd, 1998; Larkum, 2013a,b]. During the recordings, the animals
87 were exposed to randomly oriented visual stimuli with both expected and unexpected statistical properties
88 (Supp. Videos 4–5). Altogether, this approach allowed us to track the responses of both individual cell bodies
89 and individual distal apical dendritic branches over multiple days (Fig. 1E), during which the animals were
90 provided with more exposure to unexpected events. The resulting data showed evidence corroborating all three
91 of the signatures of hierarchical predictive learning above, supporting the hypothesis that the visual cortex learns
92 from unexpected events using a hierarchical model. Moreover, we observed interesting differences between the
93 distal apical dendrites and somata. Whereas somatic compartments showed a decrease in differential sensitivity
94 to expected versus unexpected visual stimuli over days, distal apical dendrites showed an increase in differential
95 sensitivity. This suggests that there may be important differences in the functional roles of the somatic and
96 distal apical compartments in hierarchical predictive learning in the neocortex.

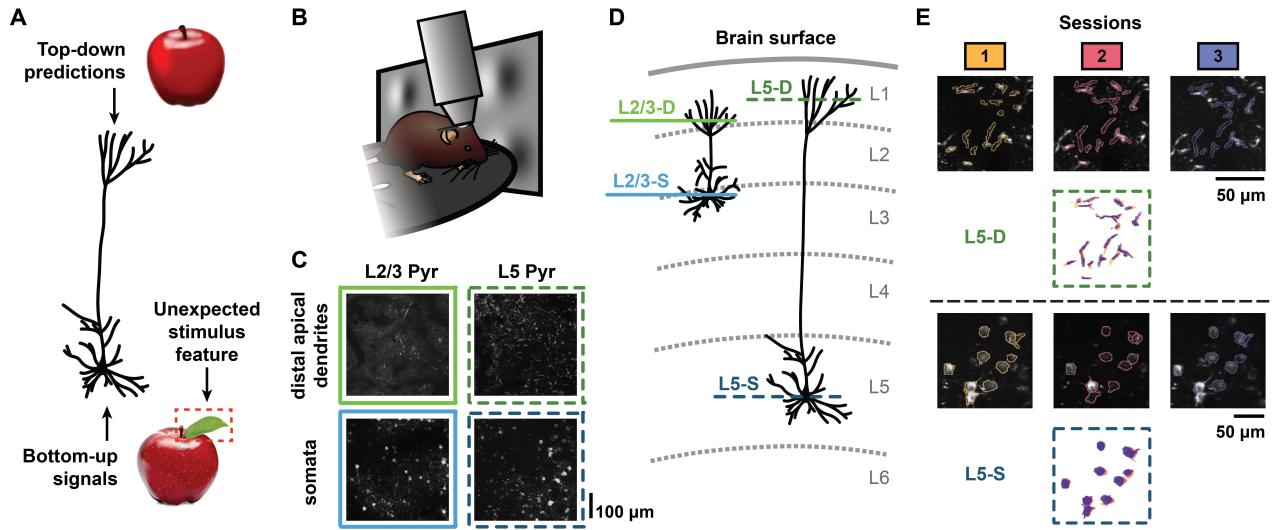


Figure 1: Illustration of experimental methods

(A) Schematic illustration of how a hierarchical predictive model might be implemented through the morphology of pyramidal neurons. The neuron receives top-down predictions (e.g., corresponding to an “expected” model of an apple, *top*) at the distal apical dendrites that are compared to bottom-up stimulus information received perisomatically (e.g., corresponding to an actual image of an apple, *bottom*). In this cartoon example, the incoming stimulus contains an unexpected feature (a leaf) not captured by the top-down predictive model.

(B) Experimental setup schematic. Awake, behaving mice were head-fixed under a two-photon microscope objective while passively viewing the stimuli. The mice were able to run freely on a rotating disc.

(C) Example maximum-projection images from two-photon recordings for each of the four imaging planes: layer 2/3 distal apical dendrites (L2/3-D), layer 5 distal apical dendrites (L5-D), layer 2/3 somata (L2/3-S), and layer 5 somata (L5-S) (2–3 mice per plane, $n = 11$ mice in total; see [Materials & Methods](#)). The corresponding recordings are shown in Supp. Video 1.

(D) Schematic illustration displaying the four imaging planes from (C) within the cortical column. The coloring and style schemes of the horizontal lines depicting the imaging planes here are used throughout all of the figures.

(E) Tracked region of interest (ROI) examples for both L5-D (*top*) and L5-S (*bottom*). Maximum-projection images for each imaging session (1, 2, or 3, as indicated), each performed on a different day, are overlaid with contours of the matched segmented ROI masks. Below the images, the matched segmented ROI masks for all three sessions are superimposed. See also Fig. S1.

97 2 Results

98 2.1 Imaging dendrite segments and cell bodies over multiple days

99 To monitor the integration of top-down and bottom-up signals by supra- and sub-granular pyramidal neurons
100 over multiple days, we performed two-photon calcium imaging in Cux2-CreERT2 mice or Rbp4-Cre_KL100 mice
101 that expressed GCaMP6f in layer 2/3 or layer 5 pyramidal neurons, respectively. We performed this imaging
102 either at layer 1 of cortex (50–75 μm depth for layer 2/3 and 20 μm depth for layer 5), thereby observing the
103 distal apical dendrites, or at the layer in which the cell bodies were located (175 μm depth for layer 2/3 and 375
104 μm depth for layer 5) (Fig. 1C–D; Supp. Video 1). This gave us four different imaging conditions: layer 2/3
105 distal apical dendrites (L2/3-D), layer 2/3 somata (L2/3-S), layer 5 distal apical dendrites (L5-D), and layer 5
106 somata (L5-S). GCaMP6f fluorescence tracks calcium influx into cells, but it should be noted that the cause
107 of calcium influx in the somatic and distal apical compartments may be different, with somatic signals largely
108 reflecting closely-spaced groups of action potentials [Huang et al., 2021] and dendritic signals largely reflecting
109 non-linear dendritic events like plateau potentials [Murayama et al., 2009]. Thus, in both cases we are tracking
110 a proxy for neural activity, but it is important to be aware that the underlying physiological cause of the signal
111 likely differs between the two compartments.

112 Imaging was performed in primary visual cortex (VisP). During the experiments, the animal’s head was

fixed in place under the microscope objective, ensuring the stability of our recordings. We extracted regions of interest (ROIs) in each imaging plane [de Vries et al., 2020; Inan et al., 2017, 2021], corresponding to individual distal apical dendrite segments or to individual cell bodies, depending on the imaging plane. Each animal went through three imaging sessions, each performed on a different day, and we used a matching algorithm to identify the same ROIs across sessions (Fig. 1E, S1).

Thanks to a very conservative quality control pipeline (see Materials & Methods), signal-to-noise ratio (SNR), $\Delta F/F$ magnitudes, and number of ROIs were stable over all three sessions in both layer 2/3 and layer 5 cell bodies and dendrites (Fig. S2). Importantly, the ROI extraction algorithm for the dendritic recordings enabled the identification of spatially discontinuous ROIs [Inan et al., 2017, 2021], reducing the risk that single dendritic compartments were split into multiple ROIs. This is supported by the observation that in both the somatic and the dendritic compartments, very few pairs of ROIs showed very high correlations in their responses (Fig. S2D). Moreover, while differences in background fluorescence levels were observable between imaging planes (Fig. 1C), these did not confound our analyses for two reasons. First, we only compared $\Delta F/F$ levels over days within each imaging plane, not between imaging planes. Second, our analysis pipeline estimated $\Delta F/F$ using a rolling baseline, so that changes in overall fluorescence would not impact our analyses (see Materials & Methods).

During these imaging sessions, we tracked the mouse's movements on a running disc (Sup. Video 2) as well as its pupil diameter with an infrared camera (Sup. Video 3). We obtained calcium imaging data for 11 mice (L2/3-D: $n = 2$, L2/3-S: $n = 3$, L5-D: $n = 3$, L5-S: $n = 3$). The full dataset is freely available online in the DANDI Archive (see Materials & Methods).

2.2 Cortical neurons respond differently to expected and unexpected stimuli

To explore the responses of cortical neurons to expected and unexpected sensory events, we designed a sequential visual stimulus inspired by previous work [Homann et al., 2017]. This stimulus had a predictable global structure, but stochastic local properties. Thanks to the predictable global structure we could randomly insert “unexpected” events, i.e., stimulus events that violated the predictable global pattern. Mice were exposed to this stimulus over multiple sessions, each occurring on different days, enabling us to observe changes in their neurons' responses to expected and unexpected sensory events.

To build a predictable global structure with some local stochasticity, we used image frames composed of randomly placed Gabor patches, assembled into five-frame sequences (*A-B-C-D-G*). Other than *G*, which was uniformly gray, each frame was defined by the locations of its Gabor patches: e.g., the locations of the Gabor patches were the same for all *A* frames for a given session, but differed between *A* and *B* frames. These Gabor patch locations were redrawn for each session, and sampled uniformly over the visual field. As a result, the locations were different in each session. Additionally, within each repeat of the sequence (*A-B-C-D-G*), the orientations of each of the Gabor patches were drawn randomly from the same distribution centered around the same mean orientation (Fig. 2A, Supp. Video 4), but the mean orientation varied from sequence to sequence. This meant that the luminance patterns at each spatial location were different for each repeat of the *A-B-C-D-G* sequence. However, because all sequences shared a global pattern wherein orientations were drawn from the same distribution across frames, knowing the orientations of the Gabors from one frame of the sequence would enable clear predictions about the orientations of the Gabors in the subsequent frames. Importantly, given these stimulus design features, the same set of images was never repeated. This reduced the risk of accommodation effects, which could cause changes in neuronal responses via mechanisms other than learning. Nonetheless, the sequences had predictable global properties that would allow an observer to form expectations about upcoming frames. Thus, the animals could learn the “rules” underlying the stimuli with increasing exposure and thereby form expectations for what should happen next. It is important to note that we cannot say with certainty whether the animals actually expected the stimulus sequences. We can, however, say that they were provided with substantial experience with which to form such expectations. For that reason, we call these *A-B-C-D-G* sequences “expected”. See Materials & Methods for a more detailed description of the stimulus properties.

To help the animals form such expectations, before the first calcium imaging session, the mice were habituated to *A-B-C-D-G* sequences over multiple sessions, each on a different day, without any violations of the predictable structure (Fig. 2B). After habituation, and during calcium imaging, the stimuli were broken up into approximately 30 blocks of randomly determined durations, each composed of repeated *A-B-C-D-G* sequences, as before. However, instead of comprising only expected sequences, each block ended with “unexpected” *A-B-C-U-G* sequences. In these sequences, the fourth frame, *D*, was replaced with an unexpected *U* frame, which had different Gabor locations and orientations. Specifically, the newly introduced *U* frames had unique random

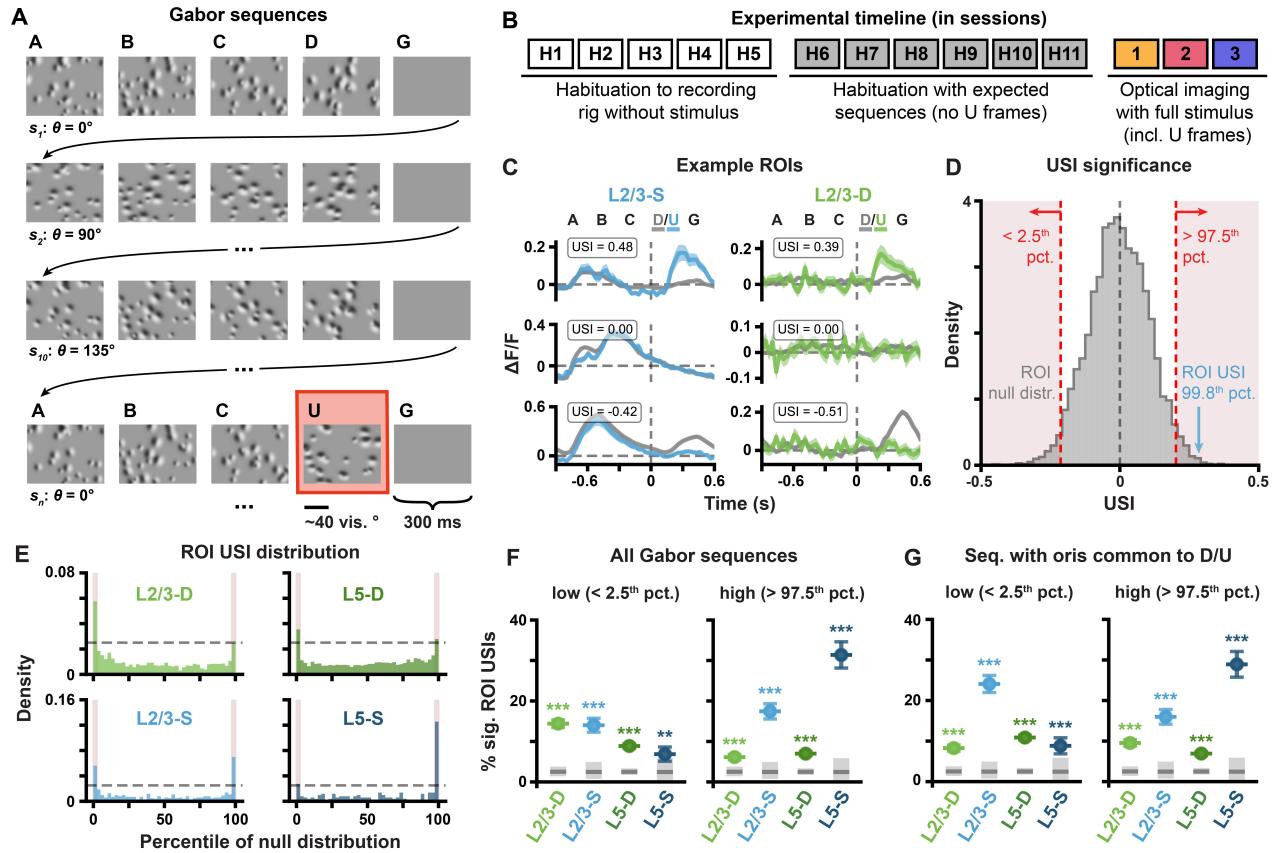


Figure 2: $\Delta F/F$ responses to unexpected stimuli reflected in the first imaging session

(A) Example Gabor sequences. Each frame lasted 300 ms. The mean orientation θ of the Gabor patches in each sequence s_i was randomly chosen from $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. An unexpected U frame, with a mean orientation rotated by 90° with respect to the other frames in the sequence, is highlighted in red. See Sec. 2.2, Materials & Methods and Supp. Video 4 for more details.

(B) Experimental timeline, showing both habituation and imaging sessions. Note that each session occurred on a different day. Optical imaging of neuronal activity was not performed during $H1-H11$.

(C) Example session 1 $\Delta F/F$ response traces for individual L2/3-S (left) and L2/3-D (right) ROIs with high (top), null (middle) or low (bottom) USIs. Mean \pm standard error of the mean (SEM) $\Delta F/F$ across Gabor sequences is plotted. Dashed vertical lines mark onset of D/U frames.

(D) Example USI null distribution for one ROI from L2/3-S in session 1, generated by shuffling $D-G$ and $U-G$ labels for the same ROI and recomputing the shuffled USIs 10^4 times. Significant regions highlighted in red, and true USI value labelled in blue.

(E) USI percentile distributions for each plane for all session 1 ROIs. Dashed horizontal lines depict null hypotheses (i.e., uniform distribution). Significant percentiles are marked with red highlights ($p < 0.05$, shuffle test as shown in (D)).

(F) Percentage \pm bootstrapped standard deviation (SD) of significant USIs for all segmented ROIs in session 1 for each plane. All sequences (any mean orientation) are included in the analysis.

(G) Same as (F), but restricted to the Gabor sequences with mean orientations shared between D and U frames $\{90^\circ, 135^\circ\}$.

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ (two-tailed, corrected).

See Table S1 for details of statistical tests and precise p-values for all comparisons.

166 locations and the orientations of the Gabor patches were resampled and shifted by 90° on average with respect
167 to the preceding *A-B-C* frames. As such, the *U* frames strongly violated potential expectations about both
168 Gabor patch locations and orientations. These “unexpected” sequences comprised approximately 7% of the
169 sequences presented to the mice during the imaging sessions.

170 ROIs in VisP exhibited clear responses to both the onset of the sequences and the final Gabor frames
171 (Fig. S3), so we compared the responses of the ROIs to the unexpected *U* frames and the expected *D* frames.
172 First, we examined the average $\Delta F/F$ signals in each of the four imaging conditions: L2/3-D, L2/3-S, L5-D, L5-
173 S. We observed that some ROIs had clearly different responses to the expected and unexpected frames (Fig. 2C,
174 *top* and *bottom* traces). To quantify the difference in responses to the expected versus the unexpected frames,
175 we calculated an “*unexpected event selectivity index*” (USI) by subtracting the mean responses to the expected
176 from the unexpected stimulus events, and scaling this value by a factor of their variances (Equation 1). We then
177 examined the USIs to see whether they indicated that the circuit treated the expected and unexpected frames
178 differently. We found that many more ROIs than would arise by chance had negative or strongly positive USIs, as
179 has been previously observed [Keller et al., 2012]. To determine chance levels, we constructed null distributions
180 non-parametrically for each ROI by shuffling the “expected” and “unexpected” labels for the stimulus frames
181 10⁴ times, each time recomputing the USI on the shuffled data (Fig. 2D; see Materials & Methods). These
182 shuffles yielded a null distribution over USI values for each ROI that reflected the null hypothesis, according to
183 which there was no difference in an ROI’s responses to expected and unexpected events. We then identified the
184 percentile of each ROI’s real USI within its own null distribution: ROI USIs below the 2.5th percentile or above
185 the 97.5th percentile were labelled as statistically significant (Fig. 2D). Across the population of ROIs, in both
186 L2/3 and L5 somata and dendrites, there were far more significant USIs than would be predicted by chance
187 (Fig. 2E–F). This effect was consistent across individual mice, with 10 of the 11 animals showing a statistically
188 significant effect (Fig. S4A). Notably, when we restricted this analysis to sequences whose mean Gabor patch
189 orientations occurred for both *D* and *U* frames, namely 90° and 135°, the USI percentages remained largely
190 the same, meaning that USI patterns did not reflect ROI preferences for specific orientations of the Gabor
191 patches (Fig. 2G). Thus, the response differences we observed were unlikely to be a result of the differences in
192 the orientations of the Gabor patches in the *D* and *U* frames. Together, these data indicate that the neurons
193 and dendrites in primary visual cortex respond systematically differently to expected and unexpected frames,
194 in line with the first observable signature of predictive learning discussed above.

195 We next wondered whether the differences in the responses to expected and unexpected frames could have
196 been driven by differences in the animals’ behavior. There is a growing body of evidence showing that responses
197 in mouse visual cortex are affected by behaviors like running and pupil dilation [Niell and Stryker, 2010; Salkoff
198 et al., 2020; Stringer et al., 2019]. Therefore, it was important to ask whether the mice altered their behavior in
199 response to the unexpected stimulus frames. If so, these behavioral differences could be reflected in the neuronal
200 responses in visual cortex, confounding our interpretation that the differences in neuronal response were due
201 to the expected versus unexpected nature of the stimulus. To test this possibility, we compared the animals’
202 running velocities and pupil dilation during the expected *D* frames and the unexpected *U* frames (Fig. 3A). We
203 found no difference in either running velocity or pupil dilation for *D* versus *U* frames (Fig. 3B), suggesting that
204 behavioral changes are not a major confound in our analyses. Altogether, these data confirm the first observable
205 signature of hierarchical predictive learning introduced above, i.e., that expected and unexpected stimuli are
206 represented differently within the neocortical microcircuit.

207 2.3 Responses to expected and unexpected stimuli evolve over days and differ 208 between the somata and distal apical dendrites

209 To probe learning, we compared the neural responses to expected and unexpected stimuli over three sessions
210 spread across multiple days. Importantly, unsupervised learning—wherein a system learns about stimuli merely
211 through exposure to them [Beaulieu and Cynader, 1990; Lotter et al., 2016; van den Oord et al., 2018; Woloszyn
212 and Sheinberg, 2012; Zylberberg et al., 2011]—is not necessarily associated with any behavioral changes. As
213 such, experimentally observing unsupervised learning requires observing changes in neural representations as
214 animals gain experience with sensory stimuli. Therefore, we examined the evolution of the neuronal responses
215 to expected and unexpected stimuli over the three different days of calcium imaging. This analysis made use of
216 our ability to track the same ROIs over each of the three imaging sessions (Fig. 1E, S1).

217 First, we examined how population-wide responses to the stimuli changed over days. In the distal apical
218 dendritic ROIs, the difference in responses to unexpected (*A-B-C-U-G*) and expected (*A-B-C-D-G*) sequences

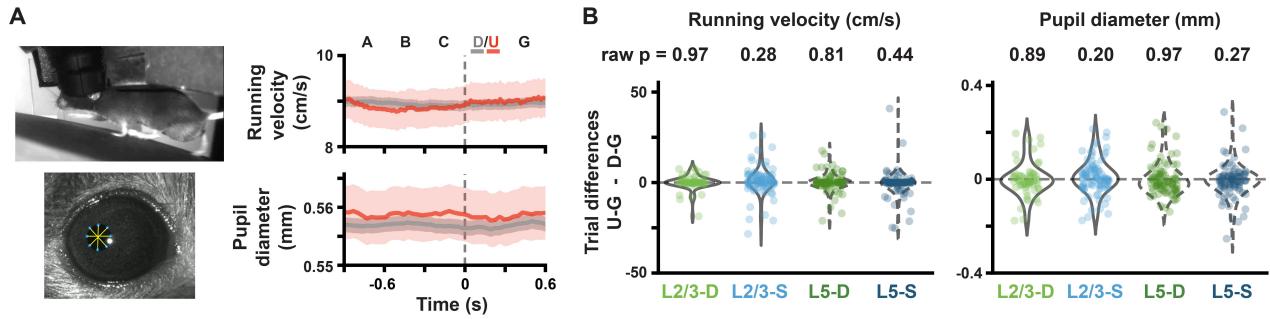


Figure 3: Absence of behavioral responses to unexpected stimuli

(A) (Left) Example frames of a mouse running (*top*), and of a mouse pupil with tracking markers (*bottom*). (Right) Running velocity and pupil diameter traces aggregated across mice (mean \pm SEM across Gabor sequences) for expected (gray) and unexpected (red) sequences. Note that the smaller SEM is due to the greater number of expected sequences, compared to unexpected ones. Dashed vertical lines mark onset of D/U frames.

(B) Block-by-block running velocity (*left*) and pupil diameter (*right*) differences between unexpected (*U-G*) and expected (*D-G*) frames. Raw two-tailed p-values (not corrected for multiple comparisons) are shown.

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ (two-tailed, corrected).

See Table S1 for details of statistical tests and precise p-values for all comparisons.

increased across days, reaching statistical significance in both L2/3 and L5 by session 3 (Fig. 4A–B, *top*). In contrast, by session 3, the response differences in the somatic ROIs, which were statistically significant in session 1 for L5-S, converged towards zero (Fig. 4A–B, *bottom*). Indeed, specifically comparing the responses to the regular sequence frames (*A-B-C*) and the unexpected frames (*U-G*), we found that the average somatic ROI responses tended to decrease for both expected and unexpected frames over time, though the effect was only statistically significant in L2/3 (Fig. 4A–C, *bottom*). In contrast, in the distal apical dendritic ROIs, we observed an increase in the average responses to the unexpected frames, but not to the regular sequence frames (Fig. 4A–C, *top*). These results indicate that the responses to the unexpected stimuli evolved differently from the responses to the regular sequence frames in these different compartments.

Importantly, there is evidence that representations in the brain can drift naturally over time, even in the absence of learning [Deitch et al., 2021; Rule et al., 2019]. As such, our above analyses left open the possibility that the changes we observed in the neural responses were not a result of unexpected event-driven learning, but were simply a result of non-specific representational drift. We think this is unlikely because we saw strong directionality to the changes in responses over days that would not be expected from random representational drift: somatic responses to unexpected events decreased towards zero, while distal apical dendritic responses increased across days (Fig. 4A–C).

Nonetheless, to further test for non-specific drift, we also examined the evolution of the responses of the same ROIs to a different, visual flow stimulus (Fig. S5A), which, based on prior work, was unlikely to drive strong expectation violations due to the fact that the visual flow was not coupled to the animals' movements [Zmarz and Keller, 2016]. In line with this previous work, we observed that although this stimulus drove changes in L2/3-S and L2/3-D, responses in L5-S and L5-D were fairly stable over sessions (Fig. S5B–C) [Jordan and Keller, 2020]. Moreover, in all compartments, the changes in responses to unexpected stimuli and USIs were smaller for the visual flow stimuli than the Gabor stimuli (S6A–E). This indicates that our observations of relatively large changes in the responses to the Gabor sequences were stimulus-specific, and hence unlikely to be caused by non-specific representational drift. Altogether, these data support the idea that VisP engages in unsupervised learning in response to unexpected events.

Given our observations of changes in the responses to the Gabor sequences at the population level, we wondered whether the same effects would be observable for the tracked ROIs. This is important because changes observed in the population-wide responses could, in principle, be driven by only a few ROIs. To test this possibility, we examined the changes over days in the responses of individual ROIs. First, we observed the same patterns as described above when we focused only on the tracked ROIs: i.e., the somatic responses tended

250 to decrease for both regular sequence frames and unexpected frames, whereas the responses to the unexpected
 251 frames increased in the distal apical dendrites (Fig. S4B).

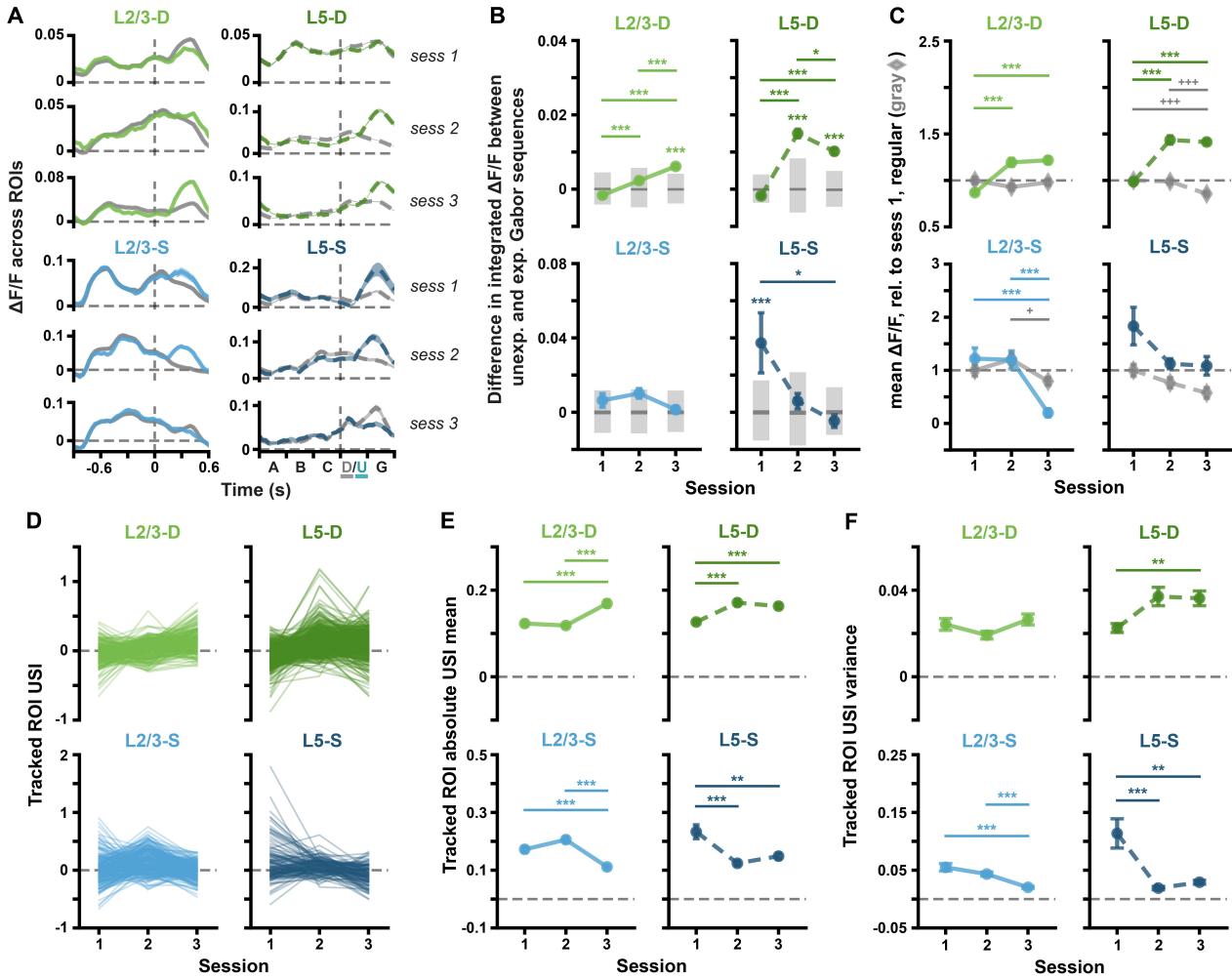


Figure 4: Unexpected Gabor sequences result in different $\Delta F/F$ and USI changes in different imaging planes.

(A) Mean (\pm SEM) across ROI mean $\Delta F/F$ responses to expected (gray, A-B-C-D-G) and unexpected (green or blue, A-B-C-U-G) Gabor sequences. Dashed vertical lines mark onset of D/U frames.

(B) Mean (\pm SEM) differences across ROIs in the mean integrated responses to expected vs. unexpected Gabor sequences, as defined in (A). Gray bars show median (dark) and adjusted 95% CIs (light) over randomly expected differences.

(C) Mean (\pm SEM) across ROIs of the mean $\Delta F/F$ responses across sequences for regular sequence frames (gray diamonds: A-B-C) and unexpected frames (green or blue circles: U-G). Responses are calculated relative to session 1 regular responses, marked by dashed horizontal lines.

(D) Gabor sequence stimulus USIs for all tracked ROIs. Each line represents a single ROI's USIs over all three sessions.

(E) Mean (\pm SEM) across the absolute values of the Gabor sequence stimulus USIs for tracked ROIs, as shown in (D).

(F) Variance (\pm bootstrapped SD) across the Gabor sequence stimulus USIs for tracked ROIs, as shown in (D).

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ (two-tailed, corrected).

+: $p < 0.05$, ++: $p < 0.01$, +++: $p < 0.001$ (two-tailed, corrected), for regular stimulus comparisons (gray) in (C).

See Table S1 for details of statistical tests and precise p-values for all comparisons.

252 Next, in order to understand how the responses to the expected *D* frames versus unexpected *U* frames
 253 evolved, we examined the USIs for the tracked ROIs over days. We found that in the somatic compartments,
 254 the USIs had converged towards zero over the three sessions (Fig. 4D-F, bottom). In contrast, in the distal apical
 255 dendritic compartments, the USIs had increased significantly over the three days. This effect was most prominent

in L2/3-D (Fig. 4D–E, *top*). These effects were generally consistent across mice (Fig. S4C). Together, these results indicate that individual ROIs altered their responses to the expected and unexpected Gabor sequence stimulus frames over multiple days, in a manner that differed between compartments receiving largely bottom-up inputs (somata) and compartments receiving largely top-down inputs (distal apical dendrites). It is important to note that if these changes were entirely random, then we would not expect to see systematic differences between the somatic and distal apical compartments, as we did here. Instead, this discovery is consistent with the second and third observable signatures of the hierarchical learning hypothesis articulated in the [Introduction](#), since we see changes over days and these differ between the distal apical dendrites and the somata.

2.4 Responses to expected and unexpected stimuli change systematically over days

Our preceding analyses showed that neurons in mouse VisP respond differently to expected and unexpected stimuli (Fig. 2), that these responses evolve over days (Fig. 4), and that there is a difference in this evolution between compartments receiving primarily bottom-up or top-down information (Fig. 4). These findings provide evidence for the three observable signatures of hierarchical predictive learning. We next sought to look for evidence that neural circuits use cell-by-cell or distal apical dendritic segment-by-segment differences in responses to expected and unexpected stimuli to guide learning.

A necessary condition for these signals to guide learning is that they contain detailed information about what was unexpected about the stimuli, i.e., information about the orientations of the Gabors. If the neural signals were to contain this information, we should be able to decode the unexpected Gabor patch orientations from the responses to the expected *D* or the unexpected *U* frames. To address this question, we trained linear logistic regression classifiers to identify the mean Gabor patch orientation from the recorded neural responses. Using a cross-validation approach with 300 random splits, we trained the classifiers for each animal and session on 75% of the data, testing them on the remaining (held-out) 25%. We found that in the somatic compartments the classifiers performed significantly above chance for session 1. This performance tended to decrease over sessions until it was at or near chance level by session 3 (Fig. 5A–B, *bottom*). In contrast, in the dendritic compartments, the performance of the decoders started above, or nearly above, chance on session 1 and then improved over sessions for the unexpected *U* frames, but not the expected *D* frames (Fig. 5A–B, *top*). Interestingly, the decoding results for *U* frames paralleled the evolution of the USIs in these compartments. Hence, the signals contain information about the nature of the unexpected orientations in a manner that reflects the extent of differences between expected and unexpected event-driven responses.

We next sought to determine whether the difference in responses to expected and unexpected stimuli was systematic across days at the level of individual ROIs. Specifically, we examined the correlation between ROI USIs in one session and the next. If the USIs do not change systematically over days, the second day's USI should resemble the first day's, plus some noise, and hence we should find positive correlations between USIs across days. Conversely, negative correlations between days are evidence of systematic changes, wherein ROIs with the largest USIs on the first day tend to have the smallest USIs on the second day, suggesting a USI-dependent learning mechanism.

To determine whether correlations were significantly different from what would be expected if there was no relationship between an ROI's USI in one session and the next, we computed null distributions for each imaging plane and session pair by shuffling 10^5 times the ROI labels within each session. Correlation values below these null distributions were interpreted as reflecting a statistically significant negative correlation between USI values across sessions for individual ROIs.

In the somatic compartments, we found no statistically significant correlations between ROI USIs in one session and the next (Fig. 5C, *bottom*). This suggests that since the overall population tendency is for somatic ROI USIs to converge towards zero over days (Fig. 4D–F, *bottom*), individual ROI USIs on one day are not linearly predictive of their values on a subsequent day. In contrast, ROI USIs in both distal apical dendritic compartments were negatively correlated from session 1 to 2 (Fig. 5C, *top*). As Fig. S7A–B shows, this reflects a tendency for the higher distal apical dendritic ROI USIs to decrease from day 1 to day 2 (*bottom right quadrants*), and for the lower ones to increase even more strongly (*top left quadrants*). ¹

¹ The previous version of this paper on *bioRxiv* showed different correlation results, which we have revised here in three ways. 1) Generally, we added more imaging sessions to our analyses to improve robustness. 2) For Fig. 5C and S6F, in order to specifically assess learning at the individual ROI level, we replaced session permutations with ROI permutations. 3) In light of these new results, we updated our interpretations of changes at the level of individual ROIs.

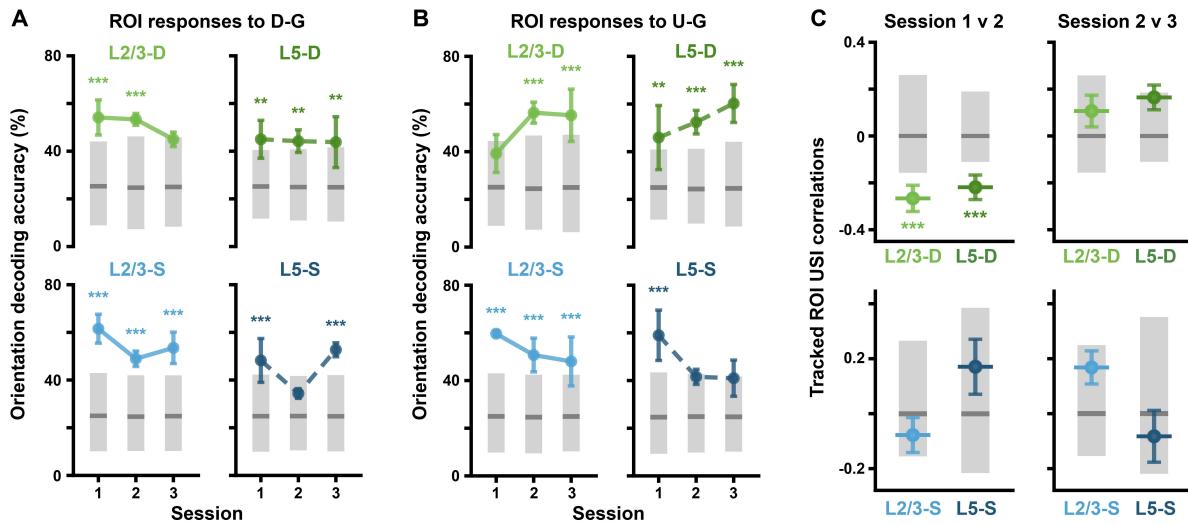


Figure 5: Unexpected Gabor sequences result in predictable $\Delta F/F$ changes in individual ROIs.

(A) Balanced accuracy (mean \pm SEM over mice) for classifiers decoding mean Gabor patch orientations from ROI activity during *D-G* frames (2–3 mice per imaging plane, 300 random cross-validation splits per mouse, per session). Gray bars show median (dark) and adjusted 95% CIs (light), computed by shuffling orientation labels.

(B) Same as (A) but with ROI activity during *U-G* frames.

(C) USI correlations (\pm bootstrapped SD) between sessions for each plane and session comparison. Gray bars show median (dark) and adjusted 95% CIs (light), computed by shuffling ROI labels.

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ (two-tailed, corrected, except for (C) where one-tailed (lower), corrected significance is reported).

See Table S1 for details of statistical tests and precise p-values for all comparisons.

In contrast, we ran this same analysis on the visual flow stimulus, where changes in representations over sessions were smaller and were only observed in L2/3-D and L2/3-S. Although none of the distal apical dendritic compartments showed USI correlations across days, ROI USIs in somatic compartments were positively correlated across days (Fig. S6F, S7C–D). Given the weak changes in mean ROI USIs across days, these correlations likely reflect a tendency for visual flow USIs to remain constant across days in the somatic compartments.

Altogether, these results demonstrate that at a dendritic segment-by-segment level, the differences between responses to expected and unexpected stimuli change systematically between days. This implicates the distinct responses to expected versus unexpected stimuli in the learning process, as anticipated under a predictive learning hypothesis.

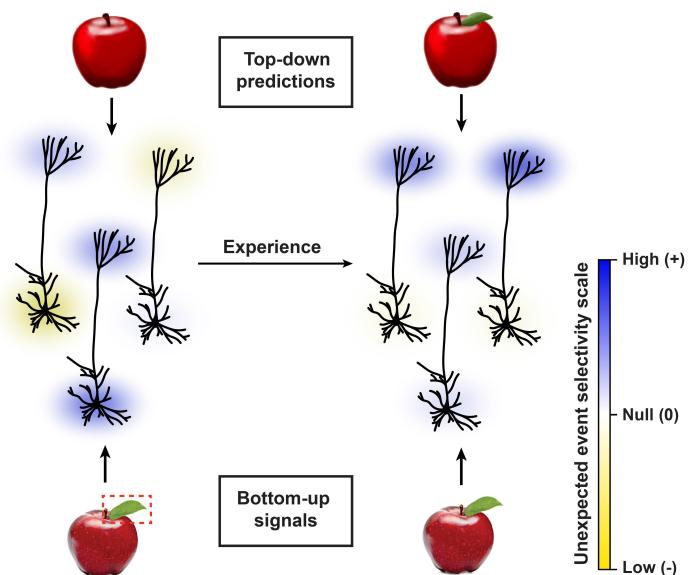
3 Discussion

In this study, we explored the question of whether the neocortex learns from unexpected stimuli. This is a central component of a broad class of theories in neuroscience and machine learning that postulate that the brain learns a hierarchical model of the world by comparing predictions about sensory stimuli to the actual stimuli received from the world. This class of theories has several observable signatures in terms of neural responses and how they should evolve over time in response to expected versus unexpected stimuli. We searched for three such observable signatures here, using chronic recordings in mouse VisP, and found evidence in support of each one. First, we observed that neurons responded differently to expected versus unexpected stimuli, which is a precondition for learning from unexpected stimuli. Second, we found that neural responses to the unexpected stimuli changed over days. In contrast, the responses to other stimuli were more stable, suggesting that the unexpected events specifically drove unsupervised learning. Third, the evolution of these responses over days differed between the distal apical dendrites (which are likely driven in large part by top-down feedback from higher-order areas [Budd, 1998; Larkum, 2013a,b]) versus the cell bodies (which are likely driven more by bottom-up sensory input [Budd, 1998; Larkum, 2013a,b]). This indicates that top-down and bottom-up signals

328 are shaped differently by the unsupervised learning process, which is a feature of learning in a hierarchical
329 model. Finally, and going beyond the three main observable signatures, we found that the sensitivity of distal
330 apical dendrites to unexpected events on one day changed systematically by the next day. This final observation
331 shows that changes in activity across days are specific to individual dendritic segments.

332 Many different forms of hierarchical unsupervised learning have been proposed. The most well-known in
333 neuroscience is probably the predictive coding model of [Rao and Ballard, 1999], along with its variations
334 [Friston and Kiebel, 2009; Spratling, 2017; Whittington and Bogacz, 2017]. But several other models in this
335 vein exist. Examples include Helmholtz machine [Dayan et al., 1995], deep belief net [Hinton and Salakhutdinov,
336 2006], Bayesian inference [Lee and Mumford, 2003], contrastive learning [Hyvärinen et al., 2019], and contrastive
337 predictive coding [van den Oord et al., 2018] models. What all of these models share is the idea that higher-
338 order association areas make predictions about incoming sensory stimuli, which then get compared to the
339 actual incoming stimuli in order to learn a model of the external world. Hence, all of these models imply the
340 experimental signatures we tested here.

341 Why do all of these models imply these same observable signatures? First, in order to learn from unexpected
342 events, there must be some available signal that distinguishes unexpected events from expected ones. Thus,
343 a key observable signature is that expected and unexpected events drive distinct responses. Second, all of
344 these models postulate that unexpected events are used to guide unsupervised learning. Thus, stimuli with
345 unexpected components should induce changes in cortical representations. Third, these models all propose that
346 higher-order areas form more abstract representations of the world, and hence the top-down signals communicate
347 something different from the bottom-up signals, which reflect incoming sensory data. Thus, learning should
348 shape these two signals differently, as they encode different aspects of the world. Finally, all of these models
349 propose that the learning algorithm utilizes the difference between expected and unexpected stimuli to shape
350 neural representations. Therefore, our data ultimately provide support for this broad class of models. Future
351 work will attempt to distinguish between the specific models within this broader class.



373 Figure 6: Learning in distal apical dendrites and somata.

374 Illustration of a conceptual model based on our data of how un-
375 expected events drive changes in the neural circuit. With ex-
376 perience, unexpected event selectivity in the somata converges toward
377 0, whereas it increases overall in the distal apical dendrites, partic-
378 ularly in dendritic segments that initially showed low selectivity.

381 computations were reflected in the different compartments of the neurons. Moreover, the finding that the distal apical
382 dendritic signals grow at a population level with exposure to the unexpected stimuli goes counter to proposals

Based on our data and previous results in the field, we propose a broad conceptual model illustrated in Fig. 6. According to this model, the brain learns an internal representation of the world in associative regions, based on which top-down predictions are provided via the distal apical dendrites to pyramidal neurons in areas like VisP. If incoming stimuli contain unexpected features, i.e., features not predicted at the distal apical dendrites (e.g., unexpected frames in Gabor sequences or an unexpected leaf on an apple, as in Fig. 6, left), pyramidal cell somatic and distal apical dendritic activity will reflect the unexpected feature or event. However, with experience, this activity triggers changes to the internal model of the world, such that it better captures the new information provided by the unexpected stimuli (e.g., by accounting for the possibility of different Gabor frames, or of apples with leaves, as in Fig. 6, right). As a result, the distal apical dendritic activity becomes more attuned to these novel forms of stimuli.

Notably, our results do not support a simple version of predictive coding wherein excitatory neurons only encode prediction errors. Although the unexpected event responses in the somata did decrease over time, in-line with encoding of errors, the responses in the dendrites increased. This suggests that different computations were reflected in the different compartments of the neurons. Moreover, the finding that the distal apical dendritic signals grow at a population level with exposure to the unexpected stimuli goes counter to proposals

383 implementing predictive coding by using the distal apical dendrite as a site for prediction error calculations
384 [Sacramento et al., 2018; Whittington and Bogacz, 2019]. More experiments with simultaneous imaging of den-
385 drites and cell bodies in different brain regions could help to clarify the distinct computational roles of these
386 neuronal compartments.

387 There are a number of limitations to this work that must be recognized. First, we were not recording
388 somata and distal apical dendrites in the same neurons. Thus, even though we saw very different evolutions
389 in the responses of the distal apical dendrites and somata to the Gabor sequence stimulus, we cannot say with
390 certainty that these differences hold within individual cells. Indeed, there is data to suggest that coupling
391 between the distal apical dendrites and the somata can sometimes be strong, particularly *in vivo* [Beaulieu-
392 Laroche et al., 2019; Francioni et al., 2019; Larkum et al., 2007]. Nonetheless, other previous work has reported
393 weak coupling between somata and distal apical dendrites [Kerlin et al., 2019; Larkum et al., 2007; Smith et al.,
394 2013], suggesting that this coupling could be context-dependent. Consistent with potential context dependence,
395 we saw clear differences in the evolution of selectivity for unexpected Gabor sequences over time between the
396 somatic and distal apical dendritic compartments, but we did not see these differences in response to the visual
397 flow stimuli (Fig. S6). Since these observations were consistent across mice (Fig. S4C), it seems likely that these
398 results would hold within individual neurons. At the same time, future work using simultaneous multi-plane
399 imaging will be critical to confirm this finding.

400 Second, though we examined the distal apical dendrites separately from the somata specifically in order to
401 identify potential differences in the processing of top-down and bottom-up inputs, an ideal experiment would
402 record simultaneously from other higher-order brain regions and their projections into visual cortex [Leinweber
403 et al., 2017; Marques et al., 2018]. This would help determine whether the signals we saw in the distal apical
404 dendrites were being calculated locally or in other regions. These experiments would be technically challenging,
405 but they are potentially feasible given recent technical advances in multi-plane mesoscope imaging.

406 Third, given the nature of our visual stimuli we were unable to measure either the classical receptive fields or
407 the orientation tuning of the neurons. As such, we cannot state with certainty whether these factors could explain
408 the differences in how individual cells responded to expected and unexpected stimuli. However, we observed
409 our results in aggregate across large populations of recorded neurons, presumably with diverse orientation
410 tuning properties and receptive fields. Thus, it is unlikely that idiosyncrasies of individual neurons' orientation
411 selectivities could account for the unexpected event responses. This assertion is supported by our finding that
412 even when we only compared responses for expected and unexpected frames with the same mean orientation we
413 still observed significant differences in the responses. Moreover, we observed significant changes in unexpected
414 event selectivity over days, whereas classic receptive fields and orientation tuning of neurons in mouse VisP are
415 known to be relatively stable over these timescales [Montijn et al., 2016].

416 Fourth, these experiments were open-loop, and thus did not incorporate any sensorimotor coupling to help
417 shape expectations. On one hand, this is a limitation given that there are a number of reports of apparent
418 sensorimotor predictions and prediction error signals in visual cortex [Keller et al., 2012; Leinweber et al., 2017;
419 Zmarz and Keller, 2016]. On the other hand, the fact that we saw evidence for learning in the open-loop
420 setting suggests that the brain is learning from sensory data alone, in addition to learning from sensorimotor
421 contingencies.

422 Fifth, and relatedly, our experiments did not incorporate any behavioral training or rewards. It could be the
423 case that the way in which the brain learns from unexpected events is different when those events are relevant
424 to motivated behaviors [Poort et al., 2015]. As such, we cannot say whether the patterns we observed would
425 carry over to task-based learning scenarios.

426 Finally, it must be recognized that different sensory stimuli, which can present different forms of unexpected
427 events, and recordings in different brain regions may produce different results. To more fully assess the hierar-
428 chical predictive learning hypothesis, future work should thoroughly explore the space of possible expected and
429 unexpected sensory stimuli and other regions of the neocortex.

430 A long-standing goal of neuroscience is to understand how our brains learn from the sensory data that
431 we receive from the world around us. Answers to this question are critical to our understanding of how we
432 build our internal models of the world, and how these govern how we interact with our surroundings. In
433 this work, we monitored changes in the responses of visual cortical neurons in mice while they learned about
434 new external stimuli, and found that these changes were consistent with a broad class of computational models,
435 namely, hierarchical predictive models. Looking forward, we anticipate that these findings could drive substantial
436 progress towards uncovering more specific models describing the brain's hierarchical predictive learning. To
437 facilitate that progress, our data and analysis software are freely available to other researchers.

4 Materials & Methods

4.1 Experimental animals and calcium imaging

The dataset used in this paper was collected as part of the Allen Institute for Brain Science's OpenScope initiative. All animal procedures were approved by the Institutional Animal Care and Use Committee (IACUC) at the Allen Institute for Brain Science. Two transgenic mouse lines (Cux2-CreERT2;Camk2a-tTA;Ai93 and Rbp4-Cre_KL100;Camk2a-tTA;Ai93) were used to drive expression of GCaMP6f in layer 2/3 and layer 5 pyramidal neurons, respectively. Mice first underwent cranial window surgery, following which they were housed in cages individually and maintained on a reverse dark-light cycle with experiments conducted during the dark phase. Mice were then habituated over two weeks to head fixation on a running disc, with the visual stimulus presentation being added the second week. Following habituation, they underwent three 70-minute optical imaging sessions within a span of three to six days, with no more than one session occurring per day. Two-photon calcium imaging was performed in the retinotopic center of VisP. Specifically, for each mouse, imaging was performed in either the cell body layer for somatic recordings (175 μm depth for layer 2/3 and 375 μm depth for layer 5) or in cortical layer 1 for distal apical dendritic recordings (50–75 μm depth for layer 2/3 and 20 μm depth for layer 5) across all optical imaging sessions. Sessions that did not meet quality control were excluded from analyses, resulting in 11 mice total (L2/3-D: $n = 2$, L2/3-S: $n = 3$, L5-D: $n = 3$, L5-S: $n = 3$) with three optical imaging sessions each. Full details on the Cre lines, surgery, habituation, and quality control are provided in [de Vries et al., 2020].

Data were collected and processed using the Allen Brain Observatory data collection and processing pipelines [de Vries et al., 2020]. Briefly, imaging was performed with Nikon A1R MP+ two-photon microscopes, and laser excitation was provided at a wavelength of 910 nm by a Ti:Sapphire laser (Chameleon Vision-Coherent). Calcium fluorescence movies were recorded at 30 Hz with resonant scanners over a 400 μm field of view with a resolution of 512 x 512 pixels (Supp. Video 1). Temporal synchronization of calcium imaging, visual stimulation, running disc movement, and infrared pupil recordings was achieved by recording all experimental clocks on a single NI PCI-6612 digital IO board at 100 kHz. Neuronal recordings were then motion corrected, and ROI masks of neuronal somata were segmented as described in [de Vries et al., 2020]. For recordings in layer 1, ROI masks of neuronal dendrites were segmented using the robust estimation algorithm developed by [Inan et al., 2017, 2021], which allows non-somatic shaped ROIs to be identified. This segmentation was run on the motion-corrected recordings, high-pass filtered spatially at 10 Hz and downsampled temporally to 15 Hz. The algorithm parameters were tuned to reject potential ROIs with a peak spatial SNR below 2.5, a temporal SNR below 5, or a spatial corruption index above 1.5, while enabling spatially discontinuous dendritic segments to be identified as part of single ROIs (Fig. S2D). Fluorescence traces for both somatic and dendritic ROIs were then extracted, neuropil-subtracted, demixed, and converted to $\Delta F/F$ traces, as described in [de Vries et al., 2020; Millman et al., 2020]. Together, neuropil subtraction and the use of a 180-second (5401 sample) sliding window to calculate rolling baseline fluorescence levels F for the $\Delta F/F$ computation ensured that the $\Delta F/F$ traces obtained were robust to potential differences in background fluorescence between mice and imaging planes. Finally, any remaining ROIs identified as being duplicates or unions, overlapping the motion border or being too noisy (defined as having a mean $\Delta F/F$ below 0 or a median $\Delta F/F$ above the midrange $\Delta F/F$, i.e., the midpoint between the minimum and maximum) were rejected. In the somatic layers, 15–224 ROIs per mouse per session were identified and retained for analysis, compared to 159–1636 ROIs in the dendritic layers. Lastly, maximum-projection images were obtained for each recording, examples of which are shown in Fig. 1C and E. Briefly, the motion corrected recordings were downsampled to ~ 4 Hz by averaging every 8 consecutive frames, following which the maximum value across downsampled frames was retained for each pixel. The resulting images were then rescaled to span the full 8-bit pixel value range (0–255). Metadata for the dataset is available on GitHub,² and the full dataset is publicly available in Neurodata Without Borders (NWB) format [Ruebel et al., 2019] in the DANDI Archive.³

4.2 Visual stimulation

During each habituation and imaging session, mice viewed the Gabor sequence stimulus, as well as a visual flow stimulus. The stimuli were presented consecutively for an equal amount of time and in random order.

² https://github.com/jeromelecoq/allen_openscope_metadata/tree/master/projects/credit_assignment

³ <https://gui.dandiarchive.org/#/dandiset/000037>

487 They appeared on a grayscreen background and were projected on a flat 24-inch monitor positioned 10 cm
488 from the right eye. The monitor was rotated and tilted to appear perpendicular to the optic axis of the eye,
489 and the stimuli were warped spatially to mimic a spherical projection screen. Whereas habituation sessions
490 increased in duration over days from 10 to 60 minutes, optical imaging sessions always lasted 70 minutes,
491 comprising 34 minutes of Gabor sequence stimulus and 17 minutes of visual flow stimulus in each direction.
492 Each stimulus period was flanked by one or 30 seconds of grayscreen for the habituation and optical imaging
493 sessions, respectively.

494 The Gabor sequence stimulus was adapted from the stimulus used in [Homann et al., 2017]. Specifically, it
495 consisted of repeating 1.5-second sequences, each comprising five consecutive frames (*A-B-C-D-G*) presented for
496 300 ms each. Whereas *G* frames were uniformly gray, frames *A*, *B*, *C*, and *D* were defined by the locations and
497 sizes of the 30 Gabor patches they each comprised. In other words, throughout a session, the locations and sizes
498 of the Gabor patches were the same for all *A* frames, but differed between *A* and *B* frames. Furthermore, these
499 locations and sizes were always resampled between mice, as well as between days, such that no two sessions
500 comprised the same Gabor sequences, even for the same mouse. The location of each Gabor patch was sampled
501 uniformly over the visual field, while its size was sampled uniformly from 10 to 20 visual degrees. Within each
502 repeat of the sequence (*A-B-C-D-G*), the orientations of each of the Gabor patches were sampled randomly
503 from a von Mises distribution with a shared mean and a kappa (dispersion parameter) of 16. This shared mean
504 orientation was randomly selected for each sequence and counterbalanced for all four orientations {0°, 45°, 90°,
505 135°}. As such, although a large range of Gabor patch orientations were viewed during a session, orientations
506 were very similar within a single sequence. “Unexpected” sequences were created by replacing *D* frames with
507 *U* frames in the sequence (*A-B-C-U-G*). *U* frames differed from *D* frames not only because they were defined
508 by a distinct set of Gabor patch sizes and locations, but also because the orientations of their Gabor patches
509 were sampled from a von Mises distribution with a mean shifted by 90° with respect to the preceding regular
510 frames (*A-B-C*), namely from {90°, 135°, 180°, 225°} (Fig. 2A, Supp. Video 4).

511 The visual flow stimulus consisted of 105 white squares moving uniformly across the screen at a velocity of
512 50 visual degrees per second, with each square being 8 by 8 visual degrees in size. The stimulus was split into
513 two consecutive periods ordered randomly, and each defined by the main direction in which the squares were
514 moving (rightward or leftward, i.e., in the nasal-to-temporal direction or vice versa, respectively). Unexpected
515 sequences, or flow violations, were created by reversing the direction of flow of a randomly selected 25% of the
516 squares for 2–4 seconds at a time, following which they resumed their motion in the main direction of flow
517 (Fig. S5A, Supp. Video 5).

518 Unexpected sequences, accounting for approximately 7% of the Gabor sequences and 5% of visual flow
519 stimulus time, *only* occurred on optical imaging days, and not on habituation days. In particular, each 70-
520 minute imaging session was broken up into approximately 30 blocks, each comprising 30–90 seconds of expected
521 sequences followed by several seconds of unexpected sequences (3–6 seconds for Gabor sequence stimulus and
522 2–4 seconds for the visual flow stimulus). All durations were sampled randomly and uniformly for each block,
523 across multiples of 1.5 seconds for the Gabor sequence stimulus and of 1 second for the visual flow stimulus.

524 The stimuli were generated using Python 2.7 [Van Rossum and Drake, 1995] custom scripts based on Psy-
525 choPy 1.82.01 [Peirce, 2009] and CamStim 0.2.4, which was developed and shared by the Allen Institute for
526 Brain Science. Code, instructions to reproduce the stimuli, and example videos are available on Github.⁴

527 4.3 Statistical analyses

528 For most analyses, mean ± standard error of the mean (SEM) is reported. In cases where the error could
529 not be directly measured over the sample, e.g., the percentage of significant ROI USIs reported in Fig. 2F, a
530 bootstrapped estimate of the error was obtained by resampling the data with replacement 10⁴ times. In these
531 cases, the standard deviation (SD) over the bootstrapped sample is plotted instead, and this is visually signaled
532 by the use of broader error caps (Fig. 2F–G, 4F, 5C, S6B, E–F).

533 Significance tests, unless otherwise indicated, were computed non-parametrically using permutation tests
534 with 10⁵ shuffles to construct null distributions, based on which confidence intervals (CIs) could be estimated.
535 Where p-values are reported, they are two-tailed (except for Fig. 5C, S6F and S7; see Sec. 4.6 Fluorescence
536 trace analysis, below), and Bonferroni-corrected for multiple comparisons to reduce the risk of Type I errors
537 (false positives). Where 95% CIs are plotted, they are equivalently adjusted using a Bonferroni correction.
538 An exception was made for Fig. 3B, which reports the relationship between the stimuli and behavioral data.

⁴ https://github.com/colleenjg/cred_assign_stimuli

539 Here, Type II errors (false negatives) were considered of greater concern, and thus we reported raw two-tailed
540 p-values in the panel itself. Details of the statistical analyses for all figures, including number of comparisons
541 and corrected p-values, are presented in Table S1.

542 4.4 Running and pupil analysis

543 Mice were allowed to run freely on a disc while head-fixed during habituation and optical imaging sessions
544 (Fig. 3A, Supp. Video 2). Running information was converted from disc rotations per running frame to cm/s.
545 The resulting velocities were median-filtered with a five-frame kernel size, and any remaining outliers, defined
546 as resulting from a single frame velocity change of at least ± 50 cm/s, were omitted from analyses.

547 To track pupil diameter during imaging sessions, an infrared LED illuminated the eye ipsilateral to the
548 monitor (right eye), allowing infrared videos to be recorded (Fig. 3A, Supp. Video 3) [Allen Institute for Brain
549 Science, 2017]. We trained a DeepLabCut model from ~ 200 manually labeled examples to automatically label
550 points around the eye, from which we estimated the pupil diameter (~ 0.01 mm per pixel conversion) [Mathis
551 et al., 2018].⁵ We omitted from analyses outlier frames, defined as resulting from a single-frame diameter change
552 of at least 0.05 mm, which usually resulted from blinking.

553 Each datapoint in Fig. 3B corresponds to the difference in the mean running velocity or pupil diameter for
554 one block between the unexpected and preceding expected Gabor sequences during session 1, with all blocks
555 being pooled across mice. We computed p-values by comparing the mean difference over all blocks for each
556 plane to a distribution of mean differences, obtained by shuffling the expected and unexpected labels 10^4 times
557 and calculating the mean difference over all blocks for each shuffle.

558 4.5 ROI tracking across sessions

559 To track ROIs across days, we employed a custom-modified version of the ROI-matching package developed
560 to track cell bodies across multiple recording days by the Allen Institute for Brain Science [de Vries et al.,
561 2020]. This pipeline implements the enhanced correlation coefficient image registration algorithm to align
562 ROI masks and the graph-theoretic blossom algorithm to optimize the separation and degree of overlap between
563 pairwise matches, as well as the number of matches across all provided sessions [Evangelidis and Psarakis, 2008].
564 This process produced highly plausible matches for the somatic ROIs; however, it provided some implausible
565 matches for the smaller and more irregularly shaped dendritic ROIs. For the dendritic ROIs, we therefore further
566 constrained the putative matches to those that overlapped by at least 10–20%. Finally, we merged results across
567 all session orderings (e.g., 1-2-3, 1-3-2, 3-1-2), eliminating any conflicting matches, i.e., non-identical matchings
568 that shared ROIs. In total, the modified matching algorithm produced ~ 100 –500 highly plausible matched
569 ROIs per plane, i.e., ~ 32 –75% of the theoretical maximum number of trackable ROIs (L2/3-D: $n = 254$, L2/3-
570 S: $n = 261$, L5-D: $n = 516$, L5-S: $n = 129$) (Fig. 1E, S1, S8).

571 4.6 Fluorescence trace analysis

572 For all results except those presented in Fig. 5A–B, S2C, and S4A, C, ROIs were pooled across all mice within an
573 imaging plane for analyses. To enable ROI pooling across mice within imaging planes, each ROI's $\Delta F/F$ trace
574 was scaled using robust standardization, i.e., by subtracting the median and then dividing by the interpercentile
575 range spanning the 5th to 95th percentile. The only additional exceptions to this are Fig. 4C, S2A-B, S4B, S6A,
576 where unscaled $\Delta F/F$ traces were used to ascertain how the $\Delta F/F$ signal itself changed across sessions.

577 Unexpected event selectivity indices (USIs) were calculated for each ROI separately using Equation 1:

$$\text{USI} = \frac{\mu_{\text{unexpected}} - \mu_{\text{expected}}}{\sqrt{\frac{1}{2} (\sigma_{\text{expected}}^2 + \sigma_{\text{unexpected}}^2)}}, \quad (1)$$

578 where the means (μ_{expected} and $\mu_{\text{unexpected}}$) and variances ($\sigma_{\text{expected}}^2$ and $\sigma_{\text{unexpected}}^2$) were calculated
579 across integrated $\Delta F/F$ responses to the expected and unexpected events, respectively. For the Gabor sequences,
580 expected events responses were defined as those spanning D-G frames, and unexpected events were defined as
581 those spanning U-G frames, with each event therefore spanning 600 ms. Indeed, G frames were included in these

⁵ https://allensdk.readthedocs.io/en/latest/allensdk.internal.brain_observatory.eye_calibration.html

582 events, as they did not introduce any new stimuli, but did consistently show persisting ROI responses to D or
583 U frames (Fig. 2C). For the visual flow stimulus, expected events were defined as the last 2 seconds of expected
584 flow before unexpected flow onset (at which point 25% of the squares reversed direction), while unexpected
585 events were defined as the first 2 seconds of unexpected flow (Fig. S5B). For each ROI, in addition to the true
586 USI, a null distribution over USIs was obtained by randomly reassigning the expected and unexpected event
587 labels to each response 10⁴ times. USIs were deemed significantly low if they lay below the 2.5th percentile, and
588 significantly high if they lay above the 97.5th percentile of their null distribution (Fig. 2D).

589 Note that for Fig. 2G, USIs were calculated using only D - G and U - G stimuli for which the mean orientations
590 were in {90°, 135°}, i.e., the orientations shared by D and U frames. For each imaging plane, the percentage
591 of significant ROI USIs was then plotted with bootstrapped SDs. Adjusted 95% CIs over chance levels were
592 estimated using the usual approximation method of the binomial CI, with the sample size corresponding to the
593 number of ROIs in the plane (Fig. 2F–G).

594 For Fig. 4A–B, ROI responses and differences in responses to full expected (A - B - C - D - G) and unexpected
595 (A - B - C - U - G) sequences were obtained by first taking the mean $\Delta F/F$ for each ROI across Gabor sequences.
596 Mean $\Delta F/F \pm$ SEM traces were then computed across ROIs and plotted for each session and imaging plane
597 (Fig. 4A). For Fig. 4B, the differences in the traces plotted in (Fig. 4A) were quantified by integrating the mean
598 $\Delta F/F$ responses over time for each ROI. Mean differences \pm SEM between expected and unexpected sequence
599 responses were then calculated across ROIs and plotted for each session and imaging plane. To further compare
600 ROI responses to regular (A - B - C) and unexpected (U - G) stimuli, for each ROI, a mean $\Delta F/F$ was calculated
601 for each set of Gabor frames, and then across sequences (Fig. 4C). The mean $\Delta F/F$ values thus obtained for
602 each ROI over a given session were then normalized by dividing by the mean $\Delta F/F$ for regular stimuli across
603 all ROIs from the same mouse in session 1. These normalized means \pm SEM over ROIs were then plotted for
604 each session and plane. Absolute fractional differences between sessions in the responses to unexpected stimuli
605 (Fig. S6B) or in USIs (Fig. S6E) were defined as

$$\left| \frac{\mu_3 - \mu_1}{\mu_1} \right|, \quad (2)$$

606 where the subscripts indicate the session over which the mean μ is computed. For Fig. S6B, μ is the mean of the
607 $\Delta F/F$ values over all ROIs for the given plane or pooled over all planes, as indicated, for unexpected sequences.
608 As in Fig. 4C, the $\Delta F/F$ values were calculated relative to the mean expected $\Delta F/F$ values on session 1 for
609 each mouse. For Fig. S6E, μ is the mean of the absolute values of the USIs for the given plane or pooled over
610 all planes for unexpected sequences. Significance tests comparing session results (Fig. 4B–C, E–F, S4B, S5C,
611 S6A, D) and those comparing Gabor sequence and visual flow stimulus results (Fig. S6B, E) were assessed by
612 permuting the session or stimulus labels, respectively, to compute adjusted 95% CIs over results expected by
613 chance.

614 For the orientation decoding analyses, linear logistic regressions were trained with an L_2 penalty on the
615 multinomial task of classifying the mean Gabor patch orientation for D - G frames {0°, 45°, 90°, 135°} or U - G
616 frames {90°, 135°, 180°, 225°}. Balanced classifier accuracy was evaluated on the test sets of 300 random
617 cross-validation 75:25 train:test splits of the dataset for each mouse. Importantly, since the D - G frame datasets
618 necessarily comprised many more examples than the U - G frame datasets (~13x), they were first downsampled
619 for each split to match the number of examples in the corresponding U - G frame datasets, thus enabling fairer
620 comparisons between D - G and U - G classification results. Input data consisted of the $\Delta F/F$ responses for all
621 ROIs together across D - G or U - G frames (600 ms). The traces were standardized as described above, but using
622 statistics drawn from the training data only. Mean balanced accuracy across dataset splits was calculated for
623 each mouse, and the mean (\pm SEM) balanced accuracy across mice was plotted for each session and plane. To
624 estimate chance accuracy, shuffled classifier performances were evaluated on 10⁵ random cross-validation dataset
625 splits for each mouse. These classifiers were trained as above, but for each split, the training set orientation
626 targets were shuffled randomly. Null distributions over mean performance were obtained by averaging classifier
627 accuracy for each split across mice, from which adjusted 95% CIs over accuracy levels expected by chance were
628 calculated for each session and plane (Fig. 5A–B).

629 Pearson correlation coefficients (Fig. 5C, S6F), and the corresponding regression slopes (Fig. S7) were calcu-
630 lated to compare ROI USIs in each imaging plane between sessions. Bootstrapped SDs over these correlations
631 for each plane were then estimated, and adjusted 95% CIs were computed by permuting the ROI labels, such
632 that tracked ROIs were no longer matched together. Here, one-tailed (lower tail) CIs were calculated to identify
633 correlations that were more negative than expected by chance.

634 5 Analysis software

635 Analyses were performed in Python 3.6 [Van Rossum and Drake, 2009] with custom scripts that are freely
636 available on GitHub,⁶ and were developed using the following packages: NumPy [Harris et al., 2020], SciPy
637 [Jones et al., 2001], Pandas [McKinney et al., 2010], Matplotlib [Hunter, 2007], Scikit-learn 0.21.1 [Pedregosa
638 et al., 2011], and the AllenSDK 1.6.0.⁷ Dendritic segmentation was run in Matlab 2019a [MATLAB, 2019]
639 using the robust estimation algorithm developed by [Inan et al., 2017, 2021]. Pupil tracking was performed
640 using DeepLabCut 2.0.5 [Mathis et al., 2018]. ROIs were matched across sessions using a custom-modified
641 version of the n-way cell matching package developed by the Allen Institute.⁸

642 6 Acknowledgements

643 The data presented herein were obtained at the Allen Brain Observatory as part of the OpenScope project, which
644 is operated by the Allen Institute for Brain Science. We thank Carol Thompson for her work coordinating the
645 OpenScope project, as well as Christof Koch and John Phillips for their continuous support of the OpenScope
646 project. We thank Wayne Wakeman for data management and support, as well as Nadezhda Dotson, Kiet
647 Ngo and Michael Taormina for their assistance in processing serial two-photon brain sections. We also thank
648 Allan Jones for providing the critical environment that enabled our large-scale team effort. We thank the Allen
649 Institute founder, Paul G. Allen, for his vision, encouragement, and support.

650 We thank Hakan Inan and Mark Schnitzer, who generously shared with us the code for their robust estimation
651 algorithm [Inan et al., 2017, 2021], and took the time to help us identify the optimal hyperparameter settings
652 for performing dendritic segmentation on the two-photon calcium imaging recordings used in this paper. We
653 also thank Daniel Denman, Stuart Trenholm, and Hubert Banville for helpful feedback on the manuscript.

654 This work was enabled by the resources provided by Compute Ontario and Compute Canada.⁹

655 7 Author Contributions

656 These authors contributed equally: CJG, JEP, JAL, as well as BAR, JZ. Experiments were designed by JZ,
657 BAR, TPL, YB. Data was collected by JAL, RA, YNB, SC, PG, IK, EL, JL, KM, CN, TVN, KN, JP, SS,
658 MTV, and AW. Data was analyzed by CJG, JEP, TMH. Supervision was provided by JAL, BAR, JZ, SC, PG,
659 AW. Manuscript was prepared by JEP, CJG, JZ, BAR, TMH.

660 8 Competing Interests

661 The authors declare no competing interests.

662 9 Funding

663 This work was supported by the Allen Institute and in part by the Falconwood Foundation. It was also supported
664 by a CIFAR Catalyst grant (JZ and BAR), Canada Research Chair grant (JZ), NSERC Discovery grants (JZ:
665 RGPIN-2019-06379. BAR: RGPIN-2014-04947), Ontario Early Researcher Award (BAR: ER17-13-242), Sloan
666 Fellowship in Neuroscience (JZ), CIFAR Azrieli Global Scholar Award (JZ), Canada CIFAR AI Chair grants
667 (BAR and YB), NSERC Canada Graduate Scholarship - Doctoral Program (CJG), and Ontario Graduate
668 Scholarship (CJG).

⁶ https://github.com/colleenjg/OpenScope_CA_Analysis

⁷ <https://github.com/AllenInstitute/AllenSDK>

⁸ https://github.com/AllenInstitute/ophys_nway_matching/tree/main/nway

⁹ www.computeontario.ca and www.computeCanada.ca

669 References

- 670 Allen Institute for Brain Science (2017). Visual coding overview. Technical report, <http://observatory.brain-map.org/visualcoding/>.
- 671 Bakhtiari, S., Mineault, P. J., Lillicrap, T., Pack, C. C., and Richards, B. A. (2021). The functional specialization
673 of visual cortex emerges from training parallel pathways with self-supervised predictive learning. In *Advances
674 in Neural Information Processing Systems*, volume 34, pages 2901–2910.
- 675 Beaulieu, C. and Cynader, M. (1990). Effect of the richness of the environment on neurons in cat visual cortex.
676 I. Receptive field properties. *Developmental Brain Research*, 53(1):71–81.
- 677 Beaulieu-Laroche, L., Toloza, E., Brown, N. J., and Harnett, M. T. (2019). Widespread and highly correlated
678 somato-dendritic activity in cortical layer 5. *Neuron*, 103(2):235–241.
- 679 Budd, J. M. (1998). Extrastriate feedback to primary visual cortex in primates: a quantitative analysis of
680 connectivity. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1400):1037–1044.
- 681 Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020). A simple framework for contrastive learning of
682 visual representations. In *International Conference on Machine Learning*, volume 37, pages 1597–1607.
- 683 Christensen, E. and Zylberberg, J. (2020). Models of the primate ventral stream that categorize and visualize
684 images. *bioRxiv:2020.02.21.958488*.
- 685 Dayan, P., Hinton, G. E., Neal, R. M., and Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*,
686 7(5):889–904.
- 687 de Vries, S. E., Lecoq, J. A., Buice, M. A., Groblewski, P. A., Ocker, G. K., Oliver, M., Feng, D., Cain,
688 N., Ledochowitsch, P., Millman, D., et al. (2020). A large-scale standardized physiological survey reveals
689 functional organization of the mouse visual cortex. *Nature Neuroscience*, 23(1):138–151.
- 690 Deitch, D., Rubin, A., and Ziv, Y. (2021). Representational drift in the mouse visual cortex. *Current Biology*,
691 31(19):4327–4339.
- 692 Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional
693 transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- 694 Evangelidis, G. D. and Psarakis, E. Z. (2008). Parametric image alignment using enhanced correlation coefficient
695 maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1858–1865.
- 696 Fiser, A., Mahringer, D., Oyibo, H. K., Petersen, A. V., Leinweber, M., and Keller, G. B. (2016). Experience-
697 dependent spatial expectations in mouse visual cortex. *Nature Neuroscience*, 19(12):1658–1664.
- 698 Francioni, V., Padamsey, Z., and Rochefort, N. L. (2019). High and asymmetric somato-dendritic coupling of
699 V1 layer 5 neurons independent of visual stimulation and locomotion. *Elife*, 8.
- 700 Friston, K. and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions
701 of the Royal Society B: Biological Sciences*, 364(1521):1211–1221.
- 702 Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009). The mismatch negativity: a review of
703 underlying mechanisms. *Clinical Neurophysiology*, 120(3):453–463.
- 704 Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B.,
705 Guo, Z., Gheshlaghi Azar, M., et al. (2020). Bootstrap your own latent - a new approach to self-supervised
706 learning. In *Advances in Neural Information Processing Systems*, volume 33.
- 707 Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E.,
708 Taylor, J., Berg, S., Smith, N. J., et al. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.
- 709 Hawkins, J. and Blakeslee, S. (2004). *On Intelligence*. Macmillan.
- 710 Higgins, I., Stringer, S., and Schnupp, J. (2017). Unsupervised learning of temporal features for word catego-
711 rization in a spiking neural network model of the auditory brain. *PLoS One*, 12(8):e0180174.

- 712 Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks.
713 *Science*, 313(5786):504–507.
- 714 Homann, J., Koay, S. A., Glidden, A. M., Tank, D. W., and Berry, M. J. (2017). Predictive coding of novel
715 versus familiar stimuli in the primary visual cortex. *bioRxiv:197608*.
- 716 Huang, L., Ledochowitsch, P., Knoblich, U., Lecoq, J., Murphy, G. J., Reid, R. C., de Vries, S. E., Koch, C.,
717 Zeng, H., Buice, M. A., et al. (2021). Relationship between simultaneously recorded spiking activity and
718 fluorescence signal in gcamp6 transgenic mice. *Elife*, 10:e51675.
- 719 Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95.
- 720 Hyvärinen, A., Sasaki, H., and Turner, R. (2019). Nonlinear ICA using auxiliary variables and generalized
721 contrastive learning. In *International Conference on Artificial Intelligence and Statistics*, volume 22, pages
722 859–868.
- 723 Inan, H., Erdogdu, M. A., and Schnitzer, M. (2017). Robust estimation of neural signals in calcium imaging.
724 In *Advances in Neural Information Processing Systems*, volume 30, pages 2901–2910.
- 725 Inan, H., Schmuckermair, C., Tasci, T., Ahanonu, B., Hernandez, O., Lecoq, J., Dinc, F., Wagner, M. J.,
726 Erdogdu, M., and Schnitzer, M. J. (2021). Fast and statistically robust cell extraction from large-scale neural
727 calcium imaging datasets. *bioRxiv:2021.03.24.436279*.
- 728 Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open source scientific tools for Python.
- 729 Jordan, R. and Keller, G. B. (2020). Opposing influence of top-down and bottom-up input on excitatory layer
730 2/3 neurons in mouse primary visual cortex. *Neuron*.
- 731 Keller, G. B., Bonhoeffer, T., and Hübener, M. (2012). Sensorimotor mismatch signals in primary visual cortex
732 of the behaving mouse. *Neuron*, 74(5):809–815.
- 733 Kerlin, A., Boaz, M., Flickinger, D., MacLennan, B. J., Dean, M. B., Davis, C., Spruston, N., and Svoboda, K.
734 (2019). Functional clustering of dendritic activity during decision-making. *Elife*, 8.
- 735 Kumaran, D. and Maguire, E. A. (2006). An unexpected sequence of events: mismatch detection in the human
736 hippocampus. *PLoS Biology*, 4(12):e424.
- 737 Larkum, M. E. (2013a). A cellular mechanism for cortical associations: an organizing principle for the cerebral
738 cortex. *Trends in Neurosciences*, 36(3):141–151.
- 739 Larkum, M. E. (2013b). The yin and yang of cortical layer 1. *Nature Neuroscience*, 16(2):114–115.
- 740 Larkum, M. E., Waters, J., Sakmann, B., and Helmchen, F. (2007). Dendritic spikes in apical dendrites of
741 neocortical layer 2/3 pyramidal neurons. *The Journal of Neuroscience*, 27(34):8999–9008.
- 742 Larochelle, H. and Hinton, G. E. (2010). Learning to combine foveal glimpses with a third-order Boltzmann
743 machine. In *Advances in Neural Information Processing Systems*, volume 23, pages 1243–1251.
- 744 Lee, T. S. and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical
745 Society of America A*, 20(7):1434–1448.
- 746 Leinweber, M., Ward, D. R., Sobczak, J. M., Attinger, A., and Keller, G. B. (2017). A sensorimotor circuit in
747 mouse cortex for visual flow predictions. *Neuron*, 95(6):1420–1432.
- 748 Lotter, W., Kreiman, G., and Cox, D. (2016). Deep predictive coding networks for video prediction and
749 unsupervised learning. *arXiv preprint arXiv:1605.08104*.
- 750 Marques, T., Nguyen, J., Fioreze, G., and Petreanu, L. (2018). The functional organization of cortical feedback
751 inputs to primary visual cortex. *Nature Neuroscience*, 21(5):757–764.
- 752 Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., and Bethge, M. (2018).
753 DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*,
754 21(9):1281–1289.

- 755 MATLAB (2019). *9.6.0.2030181 (R2019a)*. The MathWorks Inc., Natick, MA.
- 756 McKinney, W. et al. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th*
757 *Python in Science Conference*, volume 445, pages 51–56. Austin, TX.
- 758 Millman, D. J., Ocker, G. K., Caldejon, S., Larkin, J. D., Lee, E. K., Luviano, J., Nayan, C., Nguyen, T. V.,
759 North, K., Seid, S., et al. (2020). VIP interneurons in mouse primary visual cortex selectively enhance
760 responses to weak but specific stimuli. *Elife*, 9:e55130.
- 761 Montijn, J. S., Meijer, G. T., Lansink, C. S., and Pennartz, C. M. (2016). Population-level neural codes are
762 robust to single-neuron variability from a multidimensional coding perspective. *Cell Reports*, 16(9):2486–2498.
- 763 Murayama, M., Pérez-Garci, E., Nevian, T., Bock, T., Senn, W., and Larkum, M. E. (2009). Dendritic encoding
764 of sensory stimuli controlled by deep cortical interneurons. *Nature*, 457(7233):1137.
- 765 Niell, C. M. and Stryker, M. P. (2010). Modulation of visual responses by behavioral state in mouse visual
766 cortex. *Neuron*, 65(4):472–479.
- 767 Orlova, N., Tsyboulski, D., Najafi, F., Seid, S., Kivikas, S., Griffin, F., Leon, A., L’Heureux, Q., North, K.,
768 Swapp, J., et al. (2020). Multiplane mesoscope reveals distinct cortical interactions following expectation
769 violations. *bioRxiv:2020.10.06.328294*.
- 770 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P.,
771 Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay,
772 E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- 773 Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2:10.
- 774 Poort, J., Khan, A. G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., Bauza, M., Sahani, M., Keller, G. B.,
775 Mrsic-Flogel, T. D., et al. (2015). Learning enhances sensory and multiple non-sensory representations in
776 primary visual cortex. *Neuron*, 86(6):1478–1490.
- 777 Press, C., Kok, P., and Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences*,
778 24(1):13–24.
- 779 Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some
780 extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87.
- 781 Ruebel, O., Tritt, A., Dichter, B., Braun, T., Cain, N., Clack, N., Davidson, T. J., Dougherty, M., Fillion-
782 Robin, J.-C., Graddis, N., et al. (2019). NWB:N 2.0: An accessible data standard for neurophysiology.
783 *bioRxiv:523035*.
- 784 Rule, M. E., O’Leary, T., and Harvey, C. D. (2019). Causes and consequences of representational drift. *Current*
785 *Opinion in Neurobiology*, 58:141–147.
- 786 Sacramento, J., Ponte Costa, R., Bengio, Y., and Senn, W. (2018). Dendritic cortical microcircuits approximate
787 the backpropagation algorithm. In *Advances in Neural Information Processing Systems*, volume 31, pages
788 8721–8732.
- 789 Salkoff, D. B., Zagha, E., McCarthy, E., and McCormick, D. A. (2020). Movement and performance explain
790 widespread cortical activity in a visual detection task. *Cerebral Cortex*, 30(1):421–437.
- 791 Smith, S. L., Smith, I. T., Branco, T., and Häusser, M. (2013). Dendritic spikes enhance stimulus selectivity in
792 cortical neurons *in vivo*. *Nature*, 503(7474):115–120.
- 793 Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, 112:92–97.
- 794 Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., and Harris, K. D. (2019). Spontaneous
795 behaviors drive multidimensional, brainwide activity. *Science*, 364(6437).
- 796 van den Oord, A., Li, Y., and Vinyals, O. (2018). Representation learning with contrastive predictive coding.
797 *arXiv preprint arXiv:1807.03748*.

- 798 Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.
- 799 Van Rossum, G. and Drake, F. L. J. (1995). *Python Reference Manual*. Centrum voor Wiskunde en Informatica
800 Amsterdam.
- 801 Wayne, G., Hung, C.-C., Amos, D., Mirza, M., Ahuja, A., Grabska-Barwinska, A., Rae, J., Mirowski, P., Leibo,
802 J. Z., Santoro, A., et al. (2018). Unsupervised predictive memory in a goal-directed agent. *arXiv preprint
arXiv:1803.10760*.
- 804 Whittington, J. C. and Bogacz, R. (2017). An approximation of the error backpropagation algorithm in a
805 predictive coding network with local Hebbian synaptic plasticity. *Neural Computation*, 29(5):1229–1262.
- 806 Whittington, J. C. and Bogacz, R. (2019). Theories of error back-propagation in the brain. *Trends in Cognitive
807 Sciences*, 23(3):235–250.
- 808 Woloszyn, L. and Sheinberg, D. L. (2012). Effects of long-term visual experience on responses of distinct classes
809 of single units in inferior temporal cortex. *Neuron*, 74(1):193–205.
- 810 Zmarz, P. and Keller, G. B. (2016). Mismatch receptive fields in mouse visual cortex. *Neuron*, 92(4):766–772.
- 811 Zylberberg, J., Murphy, J. T., and DeWeese, M. R. (2011). A sparse coding model with synaptically local
812 plasticity and spiking neurons can account for the diverse shapes of V1 simple cell receptive fields. *PLoS
Computational Biology*, 7(10):e1002250.

814 10 Supplemental Info

815 10.1 Supplemental videos

816 Supp. Video 1: Sample two-photon recordings for each imaging plane

817 This video shows sample two-photon calcium imaging recordings for each of the four imaging planes (L2/3-D,
818 L2/3-S, L5-D, L5-S), which match the maximum-projection images in Fig. 1C. Each recording is from a different
819 mouse and played at 8x the original recording speed.

820

821 Supp. Video 2: Sample of a running recording

822 This video shows a sample recording of a mouse running on a disc, during stimulus presentation, on an optical
823 imaging day.

824

825 Supp. Video 3: Sample of an annotated pupil recording

826 This video shows a sample recording of a mouse pupil, during stimulus presentation, on an optical imaging day.
827 The right pupil, ipsilateral to the stimulus presentation screen, is shown. It is annotated with tracking markers
828 which are inferred by the DeepLabCut model and used to measure pupil diameter. Specifically, the small, filled
829 blue dots mark the 8 tracked pupil poles, and the yellow ellipse marks the elliptical pupil shape inferred from
830 the tracked poles.

831

832 Supp. Video 4: Gabor sequence stimulus example

833 This video shows example expected and unexpected sequences for the Gabor sequence stimulus. As described
834 in the [Materials & Methods](#), each frame lasts 300 ms, resulting in 1.5-second sequences. Within each expected
835 sequence (*A-B-C-D-G*), all Gabor patches share a mean orientation, sampled from {0°, 45°, 90°, 135°}. Within
836 each unexpected sequence (*A-B-C-U-G*), *U* frame Gabor patches are shifted by 90° with respect to the rest
837 of the sequence's mean Gabor patch orientation. In this example video, each frame is labelled at the bottom
838 right. Additionally, expected sequences are signalled with a green circle, and unexpected sequences, with a red
839 circle. None of these annotations appeared during the actual experiments, when the animals viewed the stimulus.

840

841 Supp. Video 5: Visual flow stimulus example

842 This video shows example expected and unexpected sequences for the visual flow stimulus. Example sequences
843 are shown for temporal to nasal (leftward), followed by nasal to temporal (rightward) main flow. As described in
844 the [Materials & Methods](#), during unexpected flow, 25% of the squares temporarily reverse their direction of flow.

845

846 10.2 Supplemental figures

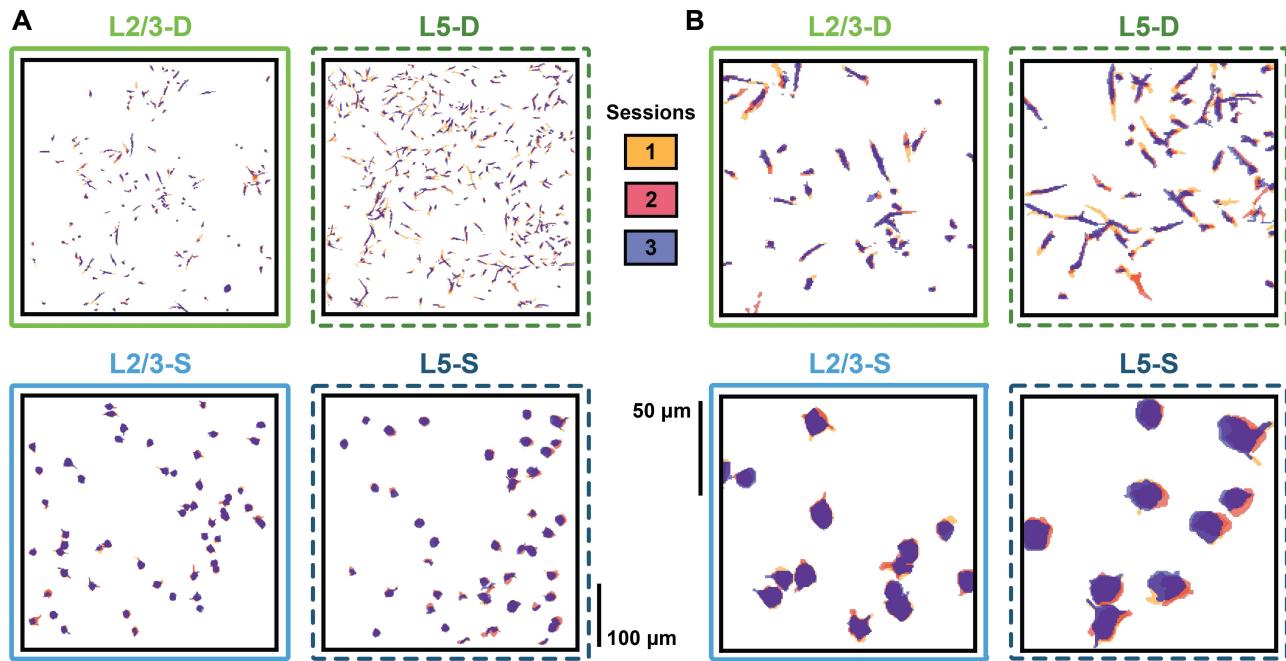


Figure S1: ROIs are successfully tracked in each plane

(A) Full field of view mask overlays for ROIs tracked across all three sessions for an example mouse in each plane.

(B) Enlarged views from (A) showing individual tracked ROI overlays for each plane.

The tracking pipeline reliably produced highly plausible ROI matches across all three sessions in each imaging plane.

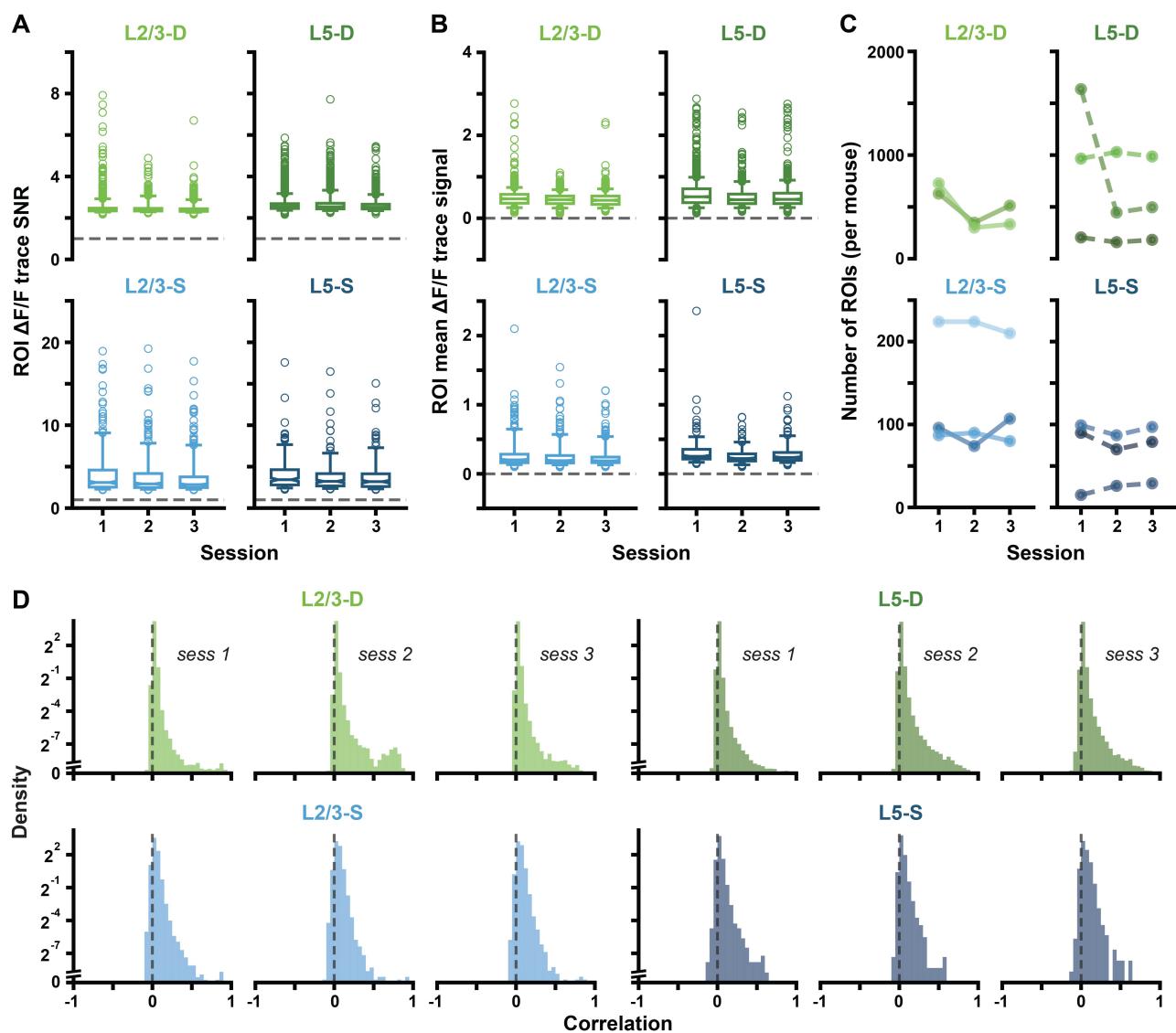


Figure S2: ROI SNR, signal and numbers are consistent across sessions

(A) $\Delta F/F$ trace SNRs for each ROI. For each session and plane, boxplots show the medians of the ROI SNR distributions, as well as the 25th to 75th percentiles, with the whiskers extending from the 5th to 95th percentiles. SNR was calculated for each ROI as follows. First, parameters (mean, SD) of a normal distribution over noisy activity were estimated based on the lower tail of the ROI's full activity distribution. The 95th percentile of the parameterized noise distribution was then defined as that ROI's noise threshold. ROI SNRs were then calculated as the ratio between their mean activity above the noise threshold (signal), and the SD of their parameterized noise distribution. Dashed horizontal lines mark 1, i.e., noise level.

SNR levels were consistent across sessions within imaging planes.

(B) Mean $\Delta F/F$ trace signal, where each datapoint corresponds to an ROI. Boxplots drawn as in (A), and signal is defined as described in (A).

Signal levels were consistent across sessions within imaging planes.

(C) The number of ROIs was generally stable across sessions for each mouse, except one in L5-D.

(D) Distributions of pairwise ROI correlations, plotted on a log scale. The log scale is linearized near 0, as signalled by the axis break. Pairwise correlations were computed over full session fluorescence traces, which were smoothed using a four-point moving average.

In all sessions, lines and planes, the vast majority of the correlation mass was concentrated near 0. The log scale reveals that the small amount of mass remaining is distributed similarly between lines and planes, largely below 0.5.

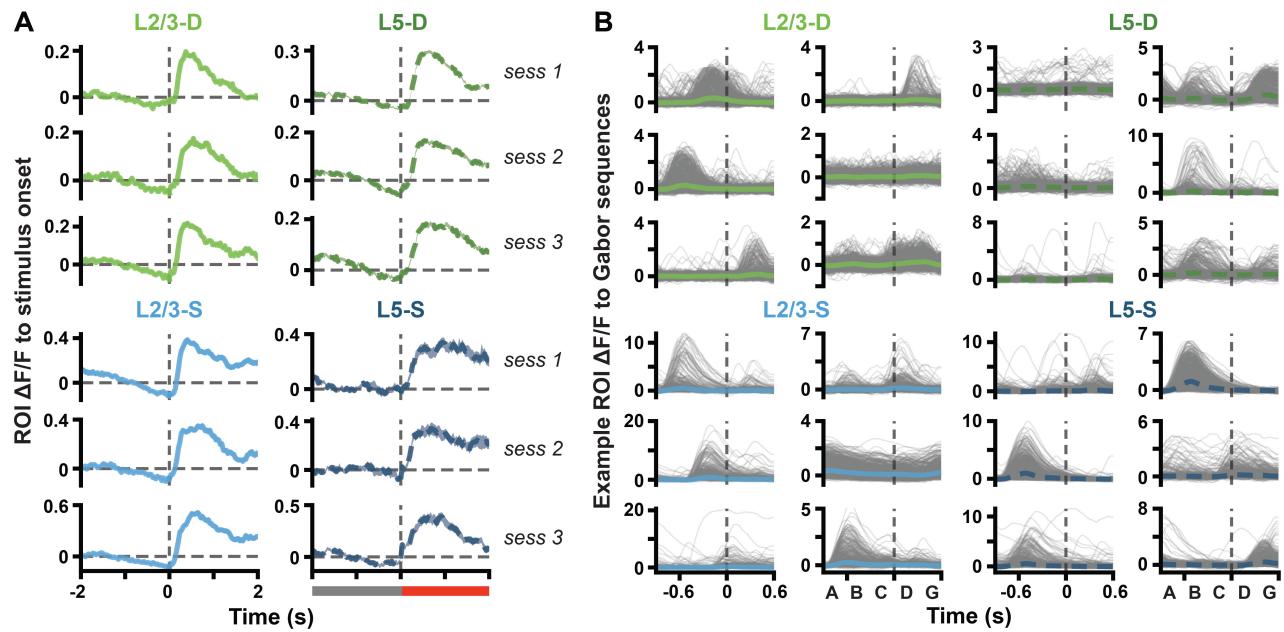


Figure S3: ROIs are responsive to stimulus onset, and to Gabor sequences

(A) Mean (\pm SEM) $\Delta F/F$ response traces across ROI mean responses to stimulus onset (Gabor sequence or visual flow) from grayscreen. Dashed vertical line at time 0 marks stimulus onset, also signalled by the gray bar becoming red (bottom of right column).

In all planes and sessions, ROI populations show clear responses to stimulus onset.

(B) $\Delta F/F$ response traces to each expected Gabor sequence (gray) for example ROIs. Mean (\pm SEM) $\Delta F/F$ responses across sequences are plotted in blue or green. Dashed vertical lines mark onset of D frames. Plotted ROIs were randomly selected from session 1 ROIs deemed consistently responsive to Gabor sequences, based on the following criteria: (1) their SNR was above the median for the session, (2) the median pairwise correlation between their individual sequence responses, as well as the SD and skew of their mean response, were each above the 75th percentile for the session. As in Fig. S2D, responses to individual sequences were smoothed using a four-point moving average, for correlation calculation and plotting, only.

In each imaging plane, numerous ROIs were found which were responsive to various components of the Gabor sequences.

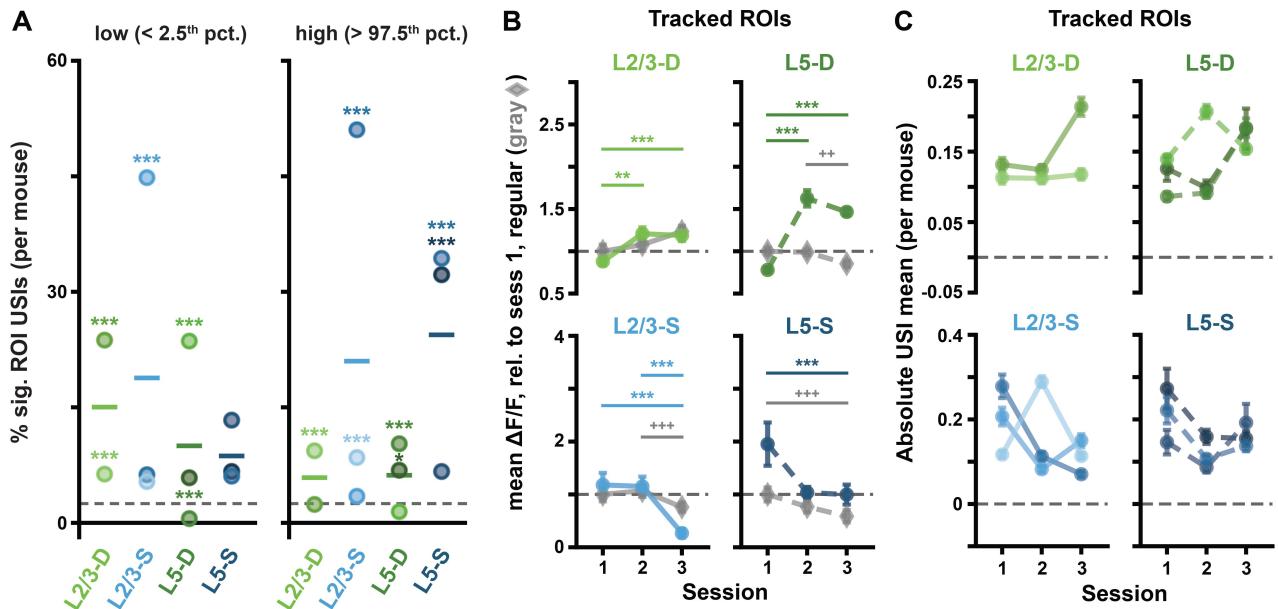


Figure S4: ROI responses to unexpected Gabor sequences are consistent in tracked ROIs and across mice

(A) Percentage of significant USIs in session 1 for each plane, where each dot corresponds to a separate mouse. Significance for each datapoint was evaluated against its own adjusted binomial CI (not shown). Lines show the pooled percentage for each plane, as plotted in Fig. 2F. Dashed horizontal lines mark the theoretical chance level (2.5%). Results are consistent with those pooled across mice, with 10 out of the 11 animals showing a higher percentage of significant ROI USIs than expected by chance in at least one tail (2F).

(B) Mean (\pm SEM) across tracked ROIs of the mean $\Delta F/F$ responses across sequences for regular sequence frames (gray diamonds: *A-B-C*) and unexpected frames (green or blue circles, *U-G*), as in Fig. 4C. Responses are calculated relative to session 1 regular responses, marked by dashed horizontal lines.

Results are consistent with the full ROI population results (Fig. 4C).

(C) Mean (\pm SEM) across the absolute values of the Gabor sequence stimulus USIs for tracked ROIs, as in Fig. 4E, but split by mouse.

Results are consistent with those pooled across mice (Fig. 4E).

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ (two-tailed, corrected).

+: $p < 0.05$, ++: $p < 0.01$, +++: $p < 0.001$ (two-tailed, corrected), for regular stimulus comparisons (gray) in (B).

See Table S1 for details of statistical tests and precise p-values for all comparisons.

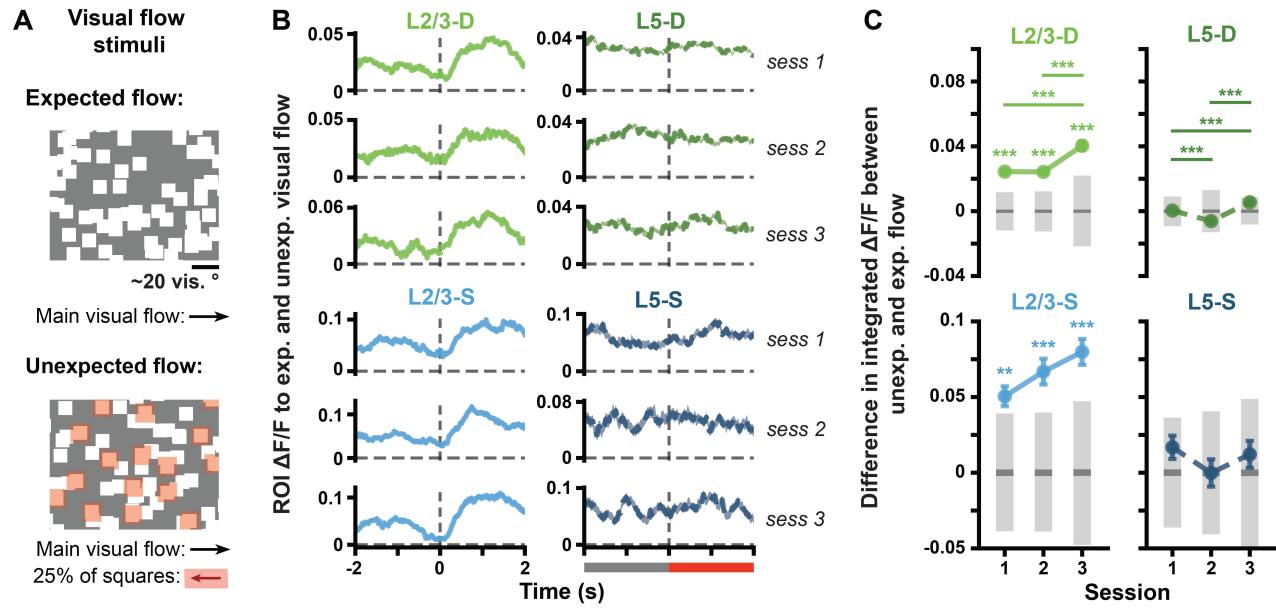


Figure S5: L2/3 ROIs respond to the onset of unexpected visual flow

(A) Visual flow stimulus. Squares moved together at the same velocity across the screen during expected flow (*top*). At random times (unexpected flow, *bottom*), 25% of the squares, highlighted here in red for illustrative purposes, reversed direction for 2–4 seconds (see Materials & Methods, and Supp. Video 5).

(B) Mean (\pm SEM) across ROI mean $\Delta F/F$ responses to visual flow sequences. Expected and unexpected visual flow sequences were defined as for the USI calculation, namely over the 2 seconds preceding unexpected visual flow onset and following its onset, respectively. Dashed vertical line at time 0 marks the onset of unexpected visual flow, also signalled by the gray bar becoming red (bottom of right column).

(C) Mean (\pm SEM) differences across ROIs in the mean integrated responses to expected vs. unexpected visual flow, as defined in (B). Gray bars show median (dark) and adjusted 95% CIs (light) over randomly expected differences. Whereas the L5 somatic and distal apical dendritic populations did not respond significantly differently to expected vs. unexpected flow, both L2/3 somatic and distal apical dendritic populations showed a significant difference in responses, which increased over days in the dendrites. These findings are consistent with recent work by [Jordan and Keller, 2020] showing that L2/3 neurons integrate visual flow mismatch information, whereas L5 neurons do not appear to do so.

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ (two-tailed, corrected).

See Table S1 for details of statistical tests and precise p-values for all comparisons.

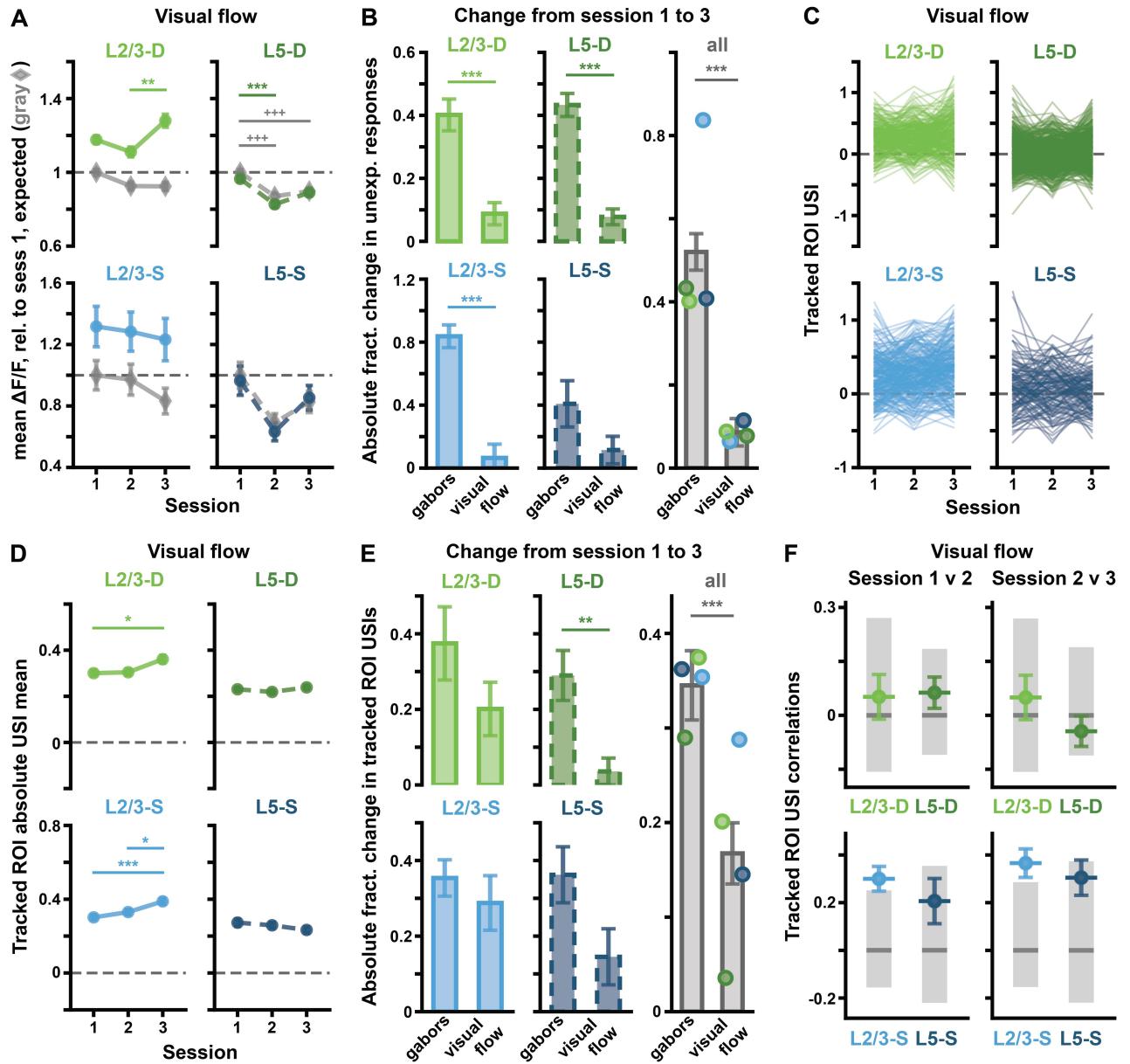


Figure S6: Unexpected visual flow sequences do not result in the same $\Delta F/F$ changes across sessions as unexpected Gabor sequences

(A) Mean (\pm SEM) across ROIs of the mean $\Delta F/F$ responses across sequences for expected flow (gray diamonds) and unexpected flow (green or blue circles), as defined in Fig. S5B. Responses are calculated relative to session 1 expected responses, marked by dashed horizontal lines. Corresponds to Fig. 4C for Gabor sequences.

(B) Absolute fractional change (\pm bootstrapped SD) in mean unexpected responses from session 1 to 3 for Gabor sequence vs. visual flow stimulus for each plane (*left* and *middle* columns), and pooled across all planes (*right* column) (see Equation 2).

In all imaging planes except L5-S, changes in ROI responses to unexpected stimulus from session 1 to 3 were significantly greater for the Gabor stimulus than for the visual flow stimulus.

(C) Visual flow stimulus USIs for all tracked ROIs. Each line represents a single ROI's USIs over all three sessions. Corresponds to Fig. 4D for Gabor sequences.

(D) Mean (\pm SEM) across the absolute values of the visual flow stimulus USIs for tracked ROIs, as shown in (C). Corresponds to Fig. 4E for Gabor sequences.

(Fig. S6 caption, cont'd)

(E) Similar to (B), but here mean (\pm bootstrapped SD) absolute fractional changes in USIs from session 1 to 3 across tracked ROIs are plotted (see Equation 2).

In L5-D and all compartments combined, changes in USIs for tracked ROIs from session 1 to 3 were significantly greater for the Gabor stimulus than for the visual flow stimulus.

(F) Correlations (\pm bootstrapped SD) for each plane and session comparison. Gray bars show median (dark) and adjusted 95% CIs (light), computed by shuffling ROI labels. Corresponds to Fig. 5C for Gabor sequences.

Unlike the Gabor sequence stimulus, only positive correlations are observed for the visual flow stimulus, and they are in the somatic compartments instead of the dendritic ones.

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ (two-tailed, corrected, except for (F) where one-tailed (lower), corrected significance is reported).

⁺: $p < 0.05$, ⁺⁺: $p < 0.01$, ⁺⁺⁺: $p < 0.001$ (two-tailed, corrected), for expected stimulus comparisons (gray) in (A).

See Table S1 for details of statistical tests and precise p-values for all comparisons.

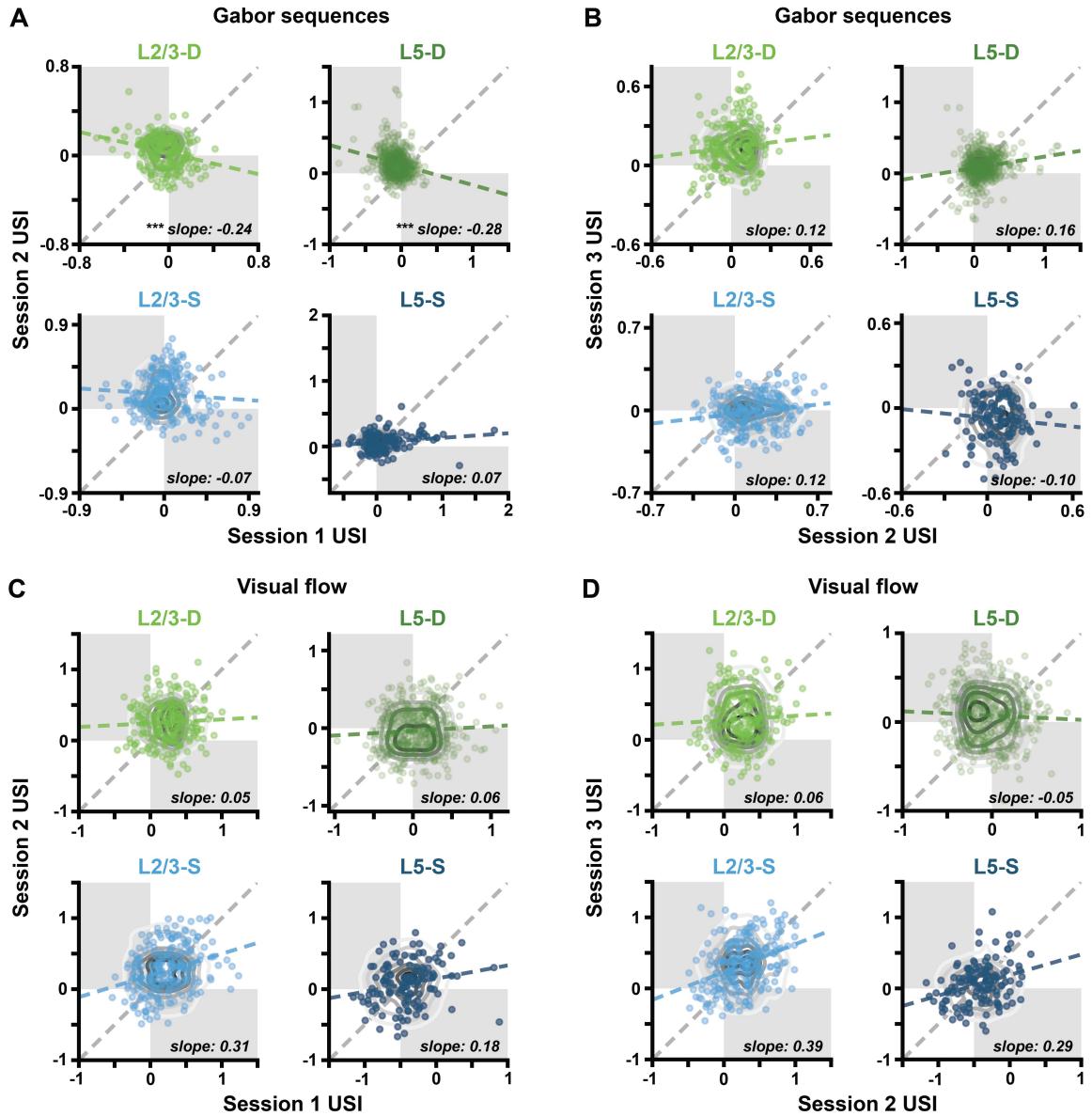


Figure S7: Gabor sequence USIs are negatively correlated between sessions 1 and 2 in the distal apical dendrites

(A) Gabor USI scatterplots showing correlations between sessions 1 and 2. Each point reflects a single tracked ROI's USIs on two sessions. Gray contour lines show null distributions, computed by shuffling ROI labels, as in Fig. 5C. The estimated regression slopes for each plane (blue or green, dashed) are plotted against the identity line (gray, dashed). Opposite quadrants are shaded in gray. Significance markers next to reported slope values correspond to the correlation significance testing results reported in Fig. 5C, and S6F.

(B) Same as in (A), but for Gabor sequence USIs in sessions 2 and 3.

(C) Same as in (A), but for visual flow USIs in sessions 1 and 2.

(D) Same as in (A), but for visual flow USIs in sessions 2 and 3.

Only L2/3-D and L5-D ROI Gabor sequence USIs show significant negative correlations. These correlations are between the USIs of sessions 1 and 2, and have slopes of -0.24 and -0.28, respectively.

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ (one-tailed (lower), corrected).

See Table S1 for details of statistical tests and precise p-values for all comparisons. Specifically, for (A–B), see the entries for Fig. 5C, and for (C–D), see the entries for Fig. S6F.

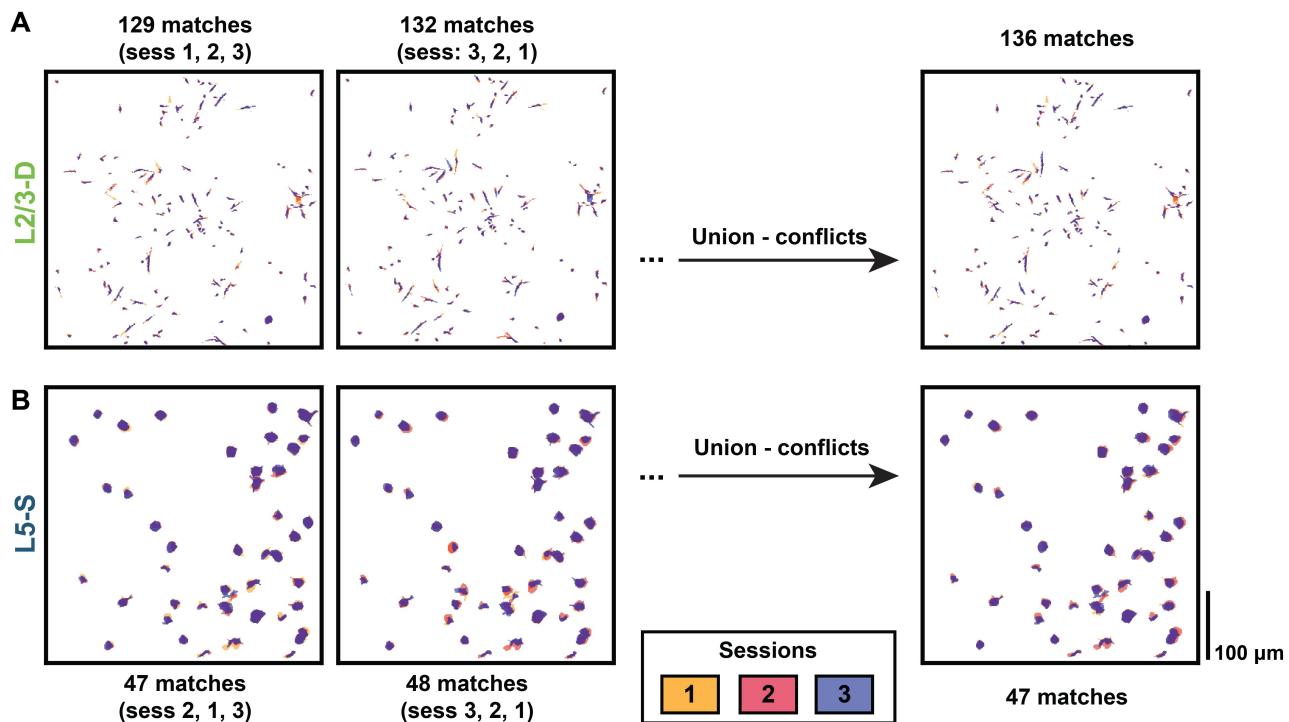


Figure S8: Dendritic ROI matches vary more with session ordering than somatic ROI matches do

(A) Example L2/3-D mouse with ROIs matched across sessions. The order in which the session images are aligned slightly affects which ROIs are matched. (*Left*) Permutation with the smallest number of matched ROIs. (*Middle*) Permutation with the largest number of matched ROIs. (*Right*) Taking the union of matches across all session permutations while removing conflicting matches (matches comprising at least one ROI that also appears in a different match) enables the quantity and quality of matched ROIs to be optimized. In this example, four pairwise matches were identified as conflicts and removed, yielding 136 final matches.

(B) Same as (A), but for a L5-S mouse. The variation in number of matched ROIs across session orderings for somata was generally far less than that for dendrites due to their larger sizes and more regular shapes. Combining matched ROIs across all permutations did nonetheless, in this example mouse, enable two of the pairwise matches to be identified as conflicts and removed, yielding 47 final matches.

847 10.3 Supplemental table

Table S1: Summary of statistical tests and results for all figures.

Fig.	Panel	Comparison	Test	Bonferroni correction	Corrected p-value	Signif.	
2	F	L2/3-D to null	binomial	8	< 0.001	$p < 0.001$	
	<i>left</i>	L2/3-S to null	null CI	comparisons	< 0.001	$p < 0.001$	
		L5-D to null			< 0.001	$p < 0.001$	
		L5-S to null			0.004	$p < 0.01$	
	F	L2/3-D to null			< 0.001	$p < 0.001$	
	<i>right</i>	L2/3-S to null			< 0.001	$p < 0.001$	
		L5-D to null			< 0.001	$p < 0.001$	
		L5-S to null			< 0.001	$p < 0.001$	
	G	L2/3-D to null	binomial	8	< 0.001	$p < 0.001$	
	<i>left</i>	L2/3-S to null	null CI	comparisons	< 0.001	$p < 0.001$	
		L5-D to null			< 0.001	$p < 0.001$	
		L5-S to null			< 0.001	$p < 0.001$	
	G	L2/3-D to null			< 0.001	$p < 0.001$	
	<i>right</i>	L2/3-S to null			< 0.001	$p < 0.001$	
		L5-D to null			< 0.001	$p < 0.001$	
		L5-S to null			< 0.001	$p < 0.001$	
3 ¹⁰	B	L2/3-D to null	10^4	4	1.000	n.s. ¹¹	
	<i>left</i>	L2/3-S to null	permutations	comparisons	1.000	n.s.	
		L5-D to null			1.000	n.s.	
		L5-S to null			1.000	n.s.	
	B	L2/3-D to null	10^4	4	1.000	n.s.	
	<i>right</i>	L2/3-S to null	permutations	comparisons	0.782	n.s.	
		L5-D to null			1.000	n.s.	
		L5-S to null			1.000	n.s.	
4	B	L2/3-D	sess. 1 to null	10^5	24	1.000	n.s.
		sess. 2 to null	permutations	comparisons	1.000	n.s.	
		sess. 3 to null			< 0.001	$p < 0.001$	
		sess. 1 vs. 2			< 0.001	$p < 0.001$	
		sess. 1 vs. 3			< 0.001	$p < 0.001$	
		sess. 2 vs. 3			< 0.001	$p < 0.001$	
		L2/3-S	sess. 1 to null			1.000	n.s.
		sess. 2 to null				0.252	n.s.
		sess. 3 to null				1.000	n.s.
		sess. 1 vs. 2				1.000	n.s.
		sess. 1 vs. 3				1.000	n.s.
		sess. 2 vs. 3				0.182	n.s.
		L5-D	sess. 1 to null			1.000	n.s.
		sess. 2 to null				< 0.001	$p < 0.001$
		sess. 3 to null				< 0.001	$p < 0.001$
		sess. 1 vs. 2				< 0.001	$p < 0.001$
		sess. 1 vs. 3				< 0.001	$p < 0.001$
		sess. 2 vs. 3				0.020	$p < 0.05$

¹⁰ In contrast to the p-values reported directly in Fig. 3, the p-values reported here are corrected for multiple comparisons.

¹¹ n.s.: not significant

Table S1: Summary of statistical tests and results for all figures. (cont'd)

Fig.	Panel	Comparison	Test	Bonferroni correction	Corrected p-value	Signif.	
(4)	(B)	L5-S	sess. 1 to null sess. 2 to null sess. 3 to null sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3	(10^5 permutations)	(24 comparisons)	< 0.001 1.000 1.000 0.755 0.011 1.000	$p < 0.001$ n.s. n.s. n.s. $p < 0.05$ n.s.
C	L2/3-D	sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3	reg. unexp.	10^5 permutations	24 comparisons	1.000 1.000 < 0.001 1.000 1.000	n.s. n.s. $p < 0.001$ n.s. n.s.
	L2/3-S	sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3	reg. unexp.			1.000 1.000 1.000 < 0.001	n.s. n.s. $p < 0.001$
	L5-D	sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3	reg. unexp.			1.000 < 0.001 < 0.001 < 0.001 1.000	n.s. $p < 0.001$ $p < 0.001$ $p < 0.001$ n.s.
	L5-S	sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3	reg. unexp.			1.000 0.814 0.118 1.000 1.000	n.s. n.s. n.s. n.s. n.s.
E	L2/3-D	sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3		10^5 permutations	12 comparisons	1.000 < 0.001 < 0.001	n.s. $p < 0.001$ $p < 0.001$
	L2/3-S	sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3				0.374 < 0.001 < 0.001	n.s. $p < 0.001$ $p < 0.001$
	L5-D	sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3				< 0.001 < 0.001 1.000	$p < 0.001$ $p < 0.001$ n.s.
	L5-S	sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3				< 0.001 0.004 0.683	$p < 0.001$ $p < 0.01$ n.s.

Table S1: Summary of statistical tests and results for all figures. (cont'd)

Fig.	Panel	Comparison	Test	Bonferroni correction	Corrected p-value	Signif.
(4)	F	L2/3-D	10^5 permutations	12	1.000	n.s.
		sess. 1 vs. 2		comparisons	1.000	n.s.
		sess. 1 vs. 3			0.647	n.s.
	L2/3-S	sess. 2 vs. 3				
		sess. 1 vs. 2			1.000	n.s.
		sess. 1 vs. 3			< 0.001	$p < 0.001$
	L5-D	sess. 2 vs. 3			< 0.001	$p < 0.001$
		sess. 1 vs. 2			0.077	n.s.
		sess. 1 vs. 3			0.008	$p < 0.01$
	L5-S	sess. 2 vs. 3			1.000	n.s.
		sess. 1 vs. 2			< 0.001	$p < 0.001$
		sess. 1 vs. 3			0.002	$p < 0.01$
		sess. 2 vs. 3			0.944	n.s.
5	A	L2/3-D	10^5 permutations	24	< 0.001	$p < 0.001$
		sess. 1 to null		comparisons	< 0.001	$p < 0.001$
		sess. 2 to null			0.086	n.s.
	L2/3-S	sess. 3 to null				
		sess. 1 to null			< 0.001	$p < 0.001$
		sess. 2 to null			< 0.001	$p < 0.001$
	L5-D	sess. 3 to null			< 0.001	$p < 0.001$
		sess. 1 to null			0.002	$p < 0.01$
		sess. 2 to null			0.006	$p < 0.01$
	L5-S	sess. 3 to null			0.008	$p < 0.01$
		sess. 1 to null			< 0.001	$p < 0.001$
		sess. 2 to null			1.000	n.s.
		sess. 3 to null			< 0.001	$p < 0.001$
B	L2/3-D	sess. 1 to null			0.525	n.s.
		sess. 2 to null			< 0.001	$p < 0.001$
		sess. 3 to null			< 0.001	$p < 0.001$
	L2/3-S	sess. 1 to null			< 0.001	$p < 0.001$
		sess. 2 to null			< 0.001	$p < 0.001$
		sess. 3 to null			< 0.001	$p < 0.001$
	L5-D	sess. 1 to null			0.003	$p < 0.01$
		sess. 2 to null			< 0.001	$p < 0.001$
		sess. 3 to null			< 0.001	$p < 0.001$
	L5-S	sess. 1 to null			< 0.001	$p < 0.001$
		sess. 2 to null			0.097	n.s.
		sess. 3 to null			0.080	n.s.
C ¹²	L2/3-D	sess. 1 v 2 to null	10^5 permutations	8	< 0.001	$p < 0.001$
		sess. 2 v 3 to null		comparison	1.000	n.s.
	L2/3-S	sess. 1 v 2 to null			0.843	n.s.
		sess. 2 v 3 to null			1.000	n.s.
	L5-D	sess. 1 v 2 to null			< 0.001	$p < 0.001$
		sess. 2 v 3 to null			1.000	n.s.

¹² One-tailed t-tests were used here (lower tail).

Table S1: Summary of statistical tests and results for all figures. (cont'd)

Fig.	Panel	Comparison	Test	Bonferroni correction	Corrected p-value	Signif.	
(5)	(C)	L5-S	sess. 1 v 2 to null sess. 2 v 3 to null	(10^5 permutations)	(8 comparisons)	1.000 1.000	n.s. n.s.
S4	A <i>left</i>	L2/3-D	~6% ~24%	binomial null CI	22 comparisons	< 0.001 < 0.001	$p < 0.001$ $p < 0.001$
		L2/3-S	~5% ~6% ~46%			0.201 0.463 < 0.001	n.s. n.s. $p < 0.001$
		L5-D	~1% ~6% ~24%			< 0.001 0.095 < 0.001	$p < 0.001$ n.s. $p < 0.001$
		L5-S	~6% ~7% ~13%			0.542 0.330 0.250	n.s. n.s. n.s.
A	<i>right</i>	L2/3-D	~2% ~10%			1.000 < 0.001	n.s. $p < 0.001$
		L2/3-S	~3% ~9% ~51%			1.000 < 0.001 < 0.001	n.s. $p < 0.001$ $p < 0.001$
		L5-D	~1% ~6% ~10%			0.767 0.010 < 0.001	n.s. $p < 0.05$ $p < 0.001$
		L5-S	~7% ~32% ~34%			1.000 < 0.001 < 0.001	n.s. $p < 0.001$ $p < 0.001$
B	L2/3-D	sess. 1 vs. 2	reg. unexp.	10^5 permutations	24 comparisons	1.000 0.003	n.s. $p < 0.01$
		sess. 1 vs. 3	reg. unexp.			0.215 < 0.001	n.s. $p < 0.001$
		sess. 2 vs. 3	reg. unexp.			1.000 1.000	n.s. n.s.
	L2/3-S	sess. 1 vs. 2	reg. unexp.			1.000 1.000	n.s. n.s.
		sess. 1 vs. 3	reg. unexp.			0.302 < 0.001	n.s. $p < 0.001$
		sess. 2 vs. 3	reg. unexp.			< 0.001 < 0.001	$p < 0.001$ $p < 0.001$
	L5-D	sess. 1 vs. 2	reg. unexp.			1.000 < 0.001	n.s. $p < 0.001$
		sess. 1 vs. 3	reg. unexp.			0.050 < 0.001	n.s. $p < 0.001$
		sess. 2 vs. 3	reg. unexp.			0.006 1.000	$p < 0.01$ n.s.

Table S1: Summary of statistical tests and results for all figures. (cont'd)

Fig.	Panel	Comparison	Test	Bonferroni correction	Corrected p-value	Signif.	
(S4)	(B)	L5-S	sess. 1 vs. 2 unexp.	10^5 permutations	(24 comparisons)	1.000 0.136	n.s. n.s.
			sess. 1 vs. 3 unexp.	reg. unexp.		< 0.001 < 0.001	$p < 0.001$ $p < 0.001$
			sess. 2 vs. 3 unexp.	reg. unexp.		1.000 1.000	n.s. n.s.
S5	C	L2/3-D	sess. 1 to null sess. 2 to null sess. 3 to null sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3	10^5 permutations	24 comparisons	< 0.001 < 0.001 < 0.001 1.000 < 0.001 < 0.001	$p < 0.001$ $p < 0.001$ $p < 0.001$ n.s. $p < 0.001$ $p < 0.001$
		L2/3-S	sess. 1 to null sess. 2 to null sess. 3 to null sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3			0.002 < 0.001 < 0.001 1.000 0.132 1.000	$p < 0.01$ $p < 0.001$ $p < 0.001$ n.s. n.s. n.s.
		L5-D	sess. 1 to null sess. 2 to null sess. 3 to null sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3			1.000 1.000 1.000 < 0.001 < 0.001 < 0.001	n.s. n.s. n.s. $p < 0.001$ $p < 0.001$ $p < 0.001$
		L5-S	sess. 1 to null sess. 2 to null sess. 3 to null sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3			1.000 1.000 1.000 1.000 1.000 1.000	n.s. n.s. n.s. n.s. n.s. n.s.
S6	A	L2/3-D	sess. 1 vs. 2 unexp.	10^5 permutations	24 comparisons	0.152 1.000	n.s. n.s.
			sess. 1 vs. 3 unexp.	exp. unexp.		0.095 0.185	n.s. n.s.
			sess. 2 vs. 3 unexp.	exp. unexp.		1.000 0.003	n.s. $p < 0.01$
		L2/3-S	sess. 1 vs. 2 unexp.	exp. unexp.		1.000 1.000	n.s. n.s.
			sess. 1 vs. 3 unexp.	exp. unexp.		1.000 1.000	n.s. n.s.
			sess. 2 vs. 3 unexp.	exp. unexp.		1.000 1.000	n.s. n.s.

Table S1: Summary of statistical tests and results for all figures. (cont'd)

Fig.	Panel	Comparison	Test	Bonferroni correction	Corrected p-value	Signif.
(S6)	(A)	L5-D	sess. 1 vs. 2 unexp.	10^5 permutations	(24 comparisons) < 0.001	< 0.001 <i>p</i> < 0.001
			sess. 1 vs. 3 unexp.			< 0.001 0.057 n.s.
			sess. 2 vs. 3 unexp.			1.000 0.344 n.s.
		L5-S	sess. 1 vs. 2 unexp.			0.072 0.067 n.s.
			sess. 1 vs. 3 unexp.			1.000 1.000 n.s.
			sess. 2 vs. 3 unexp.			1.000 0.621 n.s.
B	L2/3-D	gabors vs. vis. flow	10^5	5	< 0.001	<i>p</i> < 0.001
	L2/3-S	gabors vs. vis. flow	permutations	comparisons	< 0.001	<i>p</i> < 0.001
	L5-D	gabors vs. vis. flow			< 0.001	<i>p</i> < 0.001
	L5-S	gabors vs. vis. flow			1.000	n.s.
	all	gabors vs. vis. flow			< 0.001	<i>p</i> < 0.001
D	L2/3-D	sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3	10^5 permutations	12 comparisons	1.000 0.034 0.147	n.s. <i>p</i> < 0.05 n.s.
	L2/3-S	sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3			1.000 < 0.001 0.026	n.s. <i>p</i> < 0.001 <i>p</i> < 0.05
	L5-D	sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3			1.000 1.000 0.588	n.s. n.s. n.s.
	L5-S	sess. 1 vs. 2 sess. 1 vs. 3 sess. 2 vs. 3			1.000 1.000 1.000	n.s. n.s. n.s.
E	L2/3-D	gabors vs. vis. flow	10^5	5	0.650	n.s.
	L2/3-S	gabors vs. vis. flow	permutations	comparisons	1.000	n.s.
	L5-D	gabors vs. vis. flow			0.005	<i>p</i> < 0.01
	L5-S	gabors vs. vis. flow			0.243	n.s.
	all	gabors vs. vis. flow			< 0.001	<i>p</i> < 0.001
F ¹³	L2/3-D	sess. 1 v 2 to null sess. 2 v 3 to null	10^5 permutations	8 comparisons	1.000 1.000	n.s. n.s.
	L2/3-S	sess. 1 v 2 to null sess. 2 v 3 to null			1.000 1.000	n.s. n.s.
	L5-D	sess. 1 v 2 to null sess. 2 v 3 to null			1.000 1.000	n.s. n.s.
	L5-S	sess. 1 v 2 to null sess. 2 v 3 to null			1.000 1.000	n.s. n.s.

¹³ One-tailed t-tests were used here (lower tail).