Learning from unexpected events in the neocortical microcircuit

Colleen J. Gillon[1,2,3 *], Jason E. Pina[4,5 *], Jérôme A. Lecoq[6 *], Ruweida Ahmed[6], Yazan Billeh[6], Shiella Caldejon[6], Peter Groblewski[6], Tim M. Henley[4,5], India Kato[6], Eric Lee[6], Jennifer Luviano[6], Kyla Mace[6], Chelsea Nayan[6], Thuyanh Nguyen[6], Kat North[6], Jed Perkins[6], Sam Seid[6], Matthew Valley[6], Ali Williford[6], Yoshua Bengio[3,7,8], Timothy P. Lillicrap[9,10], Blake A. Richards[3,8,11,12 #], and Joel Zylberberg[4,5,8,13 #]

[1]Department of Biological Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada
[2]Department of Cell & Systems Biology, University of Toronto, Toronto, Ontario, Canada
[3]Mila, Montréal, Québec, Canada
[4]Department of Physics and Astronomy, York University, Toronto, Ontario, Canada
[5]Centre for Vision Research, York University, Toronto, Ontario, Canada
[6]Allen Institute for Brain Science, Seattle, WA, USA
[7]Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Québec, Canada
[8]Learning in Machines and Brains Program, Canadian Institute for Advanced Research, Toronto, Ontario, Canada
[9]DeepMind, Inc., London, UK
[10]Centre for Computation, Mathematics and Physics in the Life Sciences and Experimental Biology, University College London, London, UK
[11]School of Computer Science, McGill University, Montréal, Québec, Canada
[12]Department of Neurology & Neurosurgery, McGill University, Montréal, Québec, Canada
[13]Vector Institute for Artificial Intelligence, Toronto, Ontario, Canada
[*, #] These authors contributed equally.

January 2021

Abstract

Scientists have long conjectured that the neocortex learns the structure of the environment in a predictive, hierarchical manner. To do so, expected, predictable features are differentiated from unexpected ones by comparing bottom-up and top-down streams of data. It is theorized that the neocortex then changes the representation of incoming stimuli, guided by differences in the responses to expected and unexpected events. Such differences in cortical responses have been observed; however, it remains unknown whether these unexpected event signals govern subsequent changes in the brain's stimulus representations, and, thus, govern learning. Here, we show that unexpected event signals predict subsequent changes in responses to expected and unexpected stimuli in individual neurons and distal apical dendrites that are tracked over a period of days. These findings were obtained by observing layer 2/3 and layer 5 pyramidal neurons in primary visual cortex of awake, behaving mice using two-photon calcium imaging. We found that many neurons in both layers 2/3 and 5 showed large differences between their responses to expected and unexpected events. These unexpected event signals also determined how the responses evolved over subsequent days, in a manner that was different between the somata and distal apical dendrites. This difference 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. Together, our results provide novel evidence that the neocortex indeed instantiates a predictive hierarchical model in which unexpected events drive learning.

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

1 Introduction

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

What would constitute direct experimental evidence for the hypothesis that the brain learns a hierarchical model, where learning is guided by unexpected sensory events? This hypothesis makes four clear predictions: (1) There should be distinct responses to expected and unexpected stimuli. If the brain does not distinguish between expected and unexpected events, there is no way to specifically learn from the *unexpected* events. (2) As the circuit learns about stimuli, the responses to both expected and unexpected stimuli should change in a long-lasting manner. These changes in stimulus responses are a necessary consequence of learning modifying the stimulus representations. (3) There should be differences between the manner in which top-down and bottom-up driven responses change during learning. This follows from the idea that a hierarchical model is being learned, since hierarchy implies a distinct role for top-down and bottom-up information. (4) The changes in neural responses to stimulation should be predictable from the differences in those neurons' responses to expected and unexpected stimuli. This is a prediction of these models because they propose that the difference between the expected and unexpected stimulus responses provides the necessary information for learning. Here, we test these four predictions in the mouse visual cortex and find evidence supporting each of them. None of these predictions are individually unique to the hypothesis that the brain implements a predictive hierarchical model. Moreover, none of them are individually sufficient for demonstrating the validity of these models, but they are all necessary. Thus, it is important to test for all four, because support for all four predictions would strongly suggest that the brain implements a hierarchical predictive model.

Previous work has provided partial evidence for some of these predictions. First, there is a very large body of work showing distinct responses to expected and unexpected stimuli in multiple species and brain regions [Garrido et al., 2009; Kumaran and Maguire, 2006; Keller et al., 2012; Zmarz and Keller, 2016; Orlova et al., 2020], thus supporting the first prediction. However, there are still significant unknowns: e.g., do such responses evolve differently in different parts of the cortical microcircuit? Second, there is some research in support of the second prediction suggesting that responses to unexpected stimuli change with exposure, but only over short time scales, such as a single imaging session [Homann et al., 2017]. Third, there are a few studies showing that top-down projections carry distinct information to sensory areas [Jordan and Keller, 2020; Orlova et al., 2020], partially supporting the third prediction. However, it remains unknown whether changes in neural responses driven by top-down versus bottom-up signals differ with learning. Finally, to our knowledge, there is no data supporting the idea that the differences in responses to expected versus unexpected stimuli predict subsequent changes in signals, and thereby representations. Thus, the goal of this paper is to fill these gaps by concretely testing the four predictions together.

To address these outstanding questions, we performed chronic two-photon calcium imaging of layer 2/3 and layer 5 pyramidal neurons at both the cell bodies and the distal apical dendrites in the primary visual cortex of awake, behaving mice over multiple days (Fig. 1B-D; Supp. Videos 1–3). These imaging planes were chosen since top-down signals largely impinge on the distal apical dendrites of the layer 2/3 and layer 5 pyramidal neurons, within cortical layer 1, while bottom-up signals largely impinge on the perisomatic compartments in deeper layers [Budd, 1998; Larkum, 2013a,b]. During the recordings, the animals were exposed to randomly oriented visual stimuli with both expected and unexpected statistical properties (Supp. Videos 4–5). Altogether, this approach allowed us to track the responses of both individual cell bodies and individual distal apical dendrite branches over multiple days (Fig. 1E), as the animals were provided with more exposure to unexpected events.

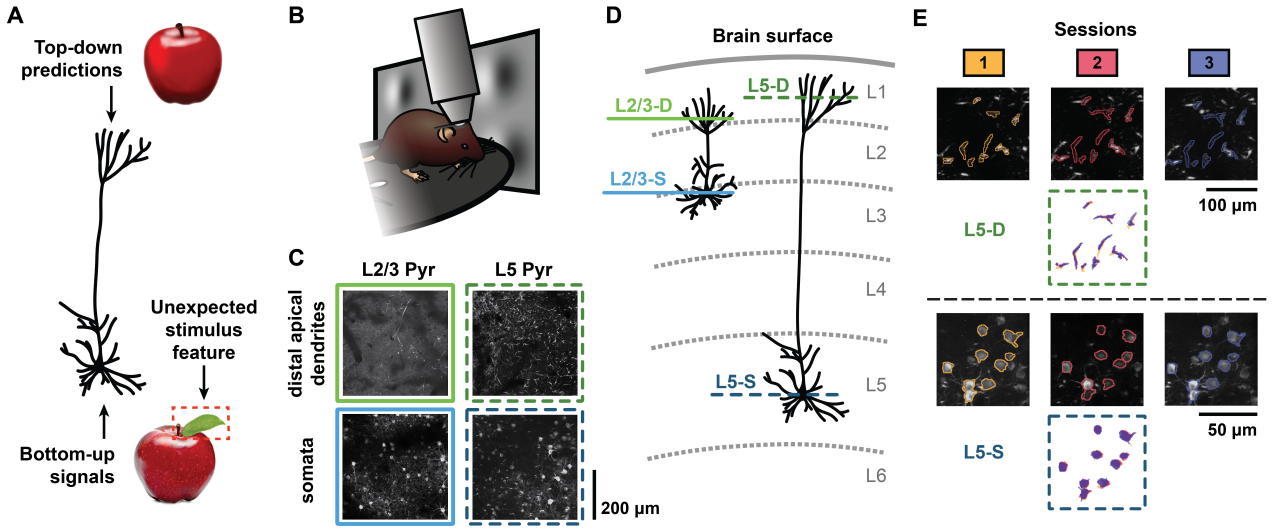

Figure 1: Neural responses to unexpected events were recorded across multiple two-photon calcium imaging sessions.

(A) Schematic illustration of how a predictive hierarchical 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 in 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 and Methods). 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 masks (contours are slightly enlarged for visibility of the ROI images). Below the images, the segmented masks across all three sessions are superimposed. See also Fig. S1, S7.

The resulting data showed evidence corroborating all four of the predictions above, supporting the hypothesis that the visual cortex uses a hierarchical model and learns from unexpected events.

2 Results

2.1 Imaging dendrite segments and cell bodies over multiple days

To monitor the integration of top-down and bottom-up signals by supra and sub–granular pyramidal neurons over multiple days, we performed two-photon calcium imaging in Cux2-CreERT2 mice or Rbp4-Cre_KL100 mice that expressed GCaMP6f in layer 2/3 or layer 5 pyramidal neurons, respectively. We performed this imaging either at layer 1 of cortex (50–75 μm depth for layer 2/3 and 20 μm depth for layer 5), thereby observing the distal apical dendrites, or at the layer in which the cell bodies were located (175 μm depth for layer 2/3 and 375 μm depth for layer 5) (Supp. Video 1). This gave us four different imaging conditions: layer 2/3 distal apical dendrites (L2/3-D), layer 2/3 somata (L2/3-S), layer 5 distal apical dendrites (L5-D), and layer 5 somata (L5-S). All 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], 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, S7). Thanks to a very conservative quality control pipeline (see Materials and 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). 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 compared responses over days within each imaging plane, not between imaging planes. Second, our analysis pipeline estimated $\Delta F/F$ using a rolling baseline, so changes in overall fluorescence would not impact our analyses (see Materials and Methods).

During these imaging sessions, we tracked the mouse's movements on a running disc (Supp. Video 2) as well as its pupil diameter with an infrared camera (Supp. 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: $n = 3$). These data are freely available online in the DANDI Archive (see Materials and 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 the Gabor patches; i.e., 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 re-drawn for each session, and sampled uniformly over the visual field. As a result, the locations were different at 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 with 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 similar 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 driven by unexpected events. 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 and Methods for a more detailed description of the stimuli 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, during calcium imaging, the stimuli were broken up into approximately 30 blocks of randomly determined durations, and composed of repeated *A-B-C-D-G* sequences, as before. However, instead of comprising only expected sequences, each block was separately by "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 locations and the orientations of the Gabor patches were resampled and shifted by 90° on average with respect to the preceding *A-B-C* frames. As such, the *U* frames strongly violated potential expectations about both Gabor patch locations and orientations. These "unexpected" sequences comprised approximately 7% of the sequences presented to the mice during the imaging sessions.

We compared the responses of the ROIs to the unexpected *U* frames and the expected *D* frames. First, we

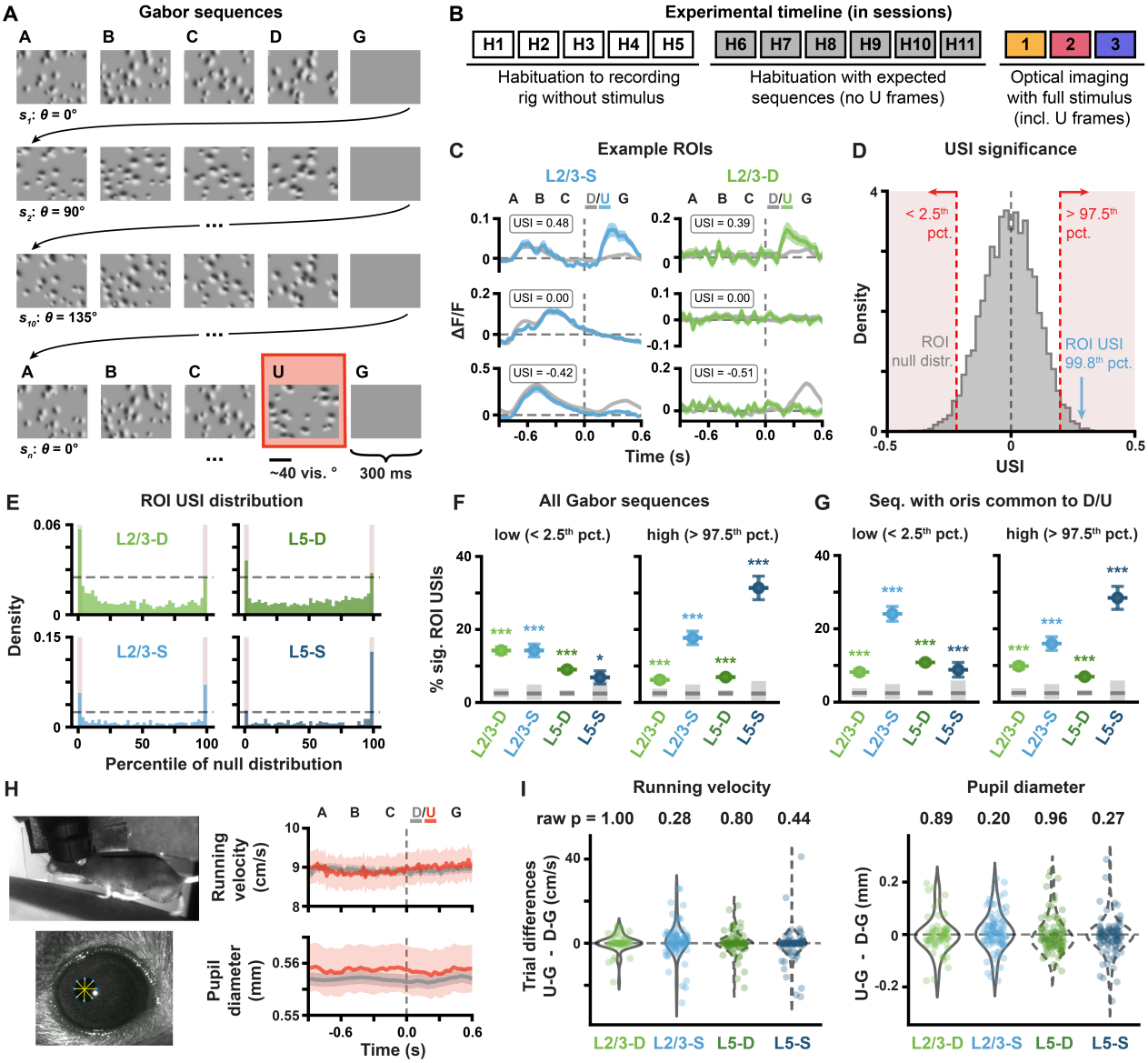

Figure 2: Unexpected stimuli reflected in $\Delta F/F$ responses, but not behavior, in the first imaging session.

(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 $\{0°, 45°, 90°, 135°\}$. 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 main text, Materials and 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 ± 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 ± 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 analysis.

(Fig. 2 caption, cont'd)
(G) Same as (F), but restricted to the Gabor sequences with mean orientations shared between D and U frames $\{90°, 135°\}$.
(H) (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 ± 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. Dashed vertical lines mark onset of D/U frames.
(I) 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.

examined the average $\Delta F/F$ signals in each of the four imaging conditions: L2/3-D, L2/3-S, L5-D, L5-S. We observed that some ROIs had clearly different responses to the expected and unexpected frames (Fig. 2C, top and bottom traces). To quantify the difference in responses to the expected versus the unexpected frames, we calculated an "***u***nexpected event ***s***electivity ***i***ndex" (USI) by subtracting the mean responses to the expected from the unexpected stimulus events, scaled by their standard deviations (Equation 1). We then examined the USIs to see whether they indicated that the circuit treated the expected and unexpected frames differently. We found that there were many more ROIs than would arise by chance with negative or strongly positive USIs, as has been previously observed [Keller et al., 2012]. To determine chance levels, we constructed null distributions non–parametrically for each ROI by shuffling the "expected" and "unexpected" labels for the stimulus frames 10^4 times, each time recomputing the USI on the shuffled data (Fig. 2D; see Materials and Methods). These shuffles yielded a null distribution over USI values for each ROI which reflected the null hypothesis according to which there was no difference in the ROI's responses to expected and unexpected events. We then identified the percentile of each ROI's real USI within its own null distribution: ROI USIs below the 2.5th percentile or above the 97.5th percentile were labelled as significant (Fig. 2D). Across the population of ROIs, in both L2/3 and L5 neurons somata and dendrites, there were far more significant USIs than would be predicted by chance (Fig. 2E–F). This effect was consistent across individual mice, with 9 of the 11 animals showing a significant effect (Fig. S4A). Notably, when we restricted this analysis to sequences whose mean Gabor patch orientations occur for both D and U frames, namely $\{90°, 135°\}$, the USI percentages remained largely the same, meaning that USI patterns did not reflect ROI preferences for specific orientations of the Gabor patches (Fig. 2G). Thus, the response differences we observed were unlikely to be a result of the differences in the orientations of the Gabor patches in the D and U frames. Together, these data indicate that the neurons and dendrites in primary visual cortex respond systematically differently to expected and unexpected frames, in line with prediction number one above.

We next wondered whether the differences in the responses to expected and unexpected frames could have been driven by differences in the animals' behavior. There is a growing body of evidence showing that responses in mouse visual cortex are affected by behaviors like running and pupil dilation [Niell and Stryker, 2010; Stringer et al., 2019; Salkoff et al., 2020]. Therefore, it was important to ask whether the mice altered their behavior in response to the unexpected stimulus frames. If so, these behavioral differences could be reflected in the neuronal responses in visual cortex, confounding our interpretation that the differences in neuronal response were due to the expected versus unexpected nature of the stimulus. To test this possibility, we compared the animals' running velocities and pupil dilation during the expected D frames and the unexpected U frames (Fig. 2H). We found no difference in either running velocity or pupil dilation for D versus U frames (Fig. 2I), suggesting that behavioral changes are not a major confound in our analyses. Altogether, these data confirm the first prediction of hierarchical learning introduced above, i.e., that expected and unexpected stimuli are represented differently within the neocortical microcircuit.

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

To probe learning, we compared the neural responses to expected and unexpected stimuli over three sessions. Importantly, unsupervised learning—wherein a system learns about stimuli merely through exposure to them [Beaulieu and Cynader, 1990; Woloszyn and Sheinberg, 2012; Zylberberg et al., 2011; van den Oord et al., 2018;

Lotter et al., 2016]—is not necessarily associated with any behavioral changes. As such, experimentally observing unsupervised learning requires observing changes in neural representations as animals gain experience with sensory stimuli. Therefore, we examined the evolution of the neuronal responses to expected and unexpected stimuli over the three different days of calcium imaging. This analysis made use of our ability to track the same ROIs over each of the three imaging sessions (Fig. 1E).

First, we examined how population-wide responses to the stimuli changed over days. In the distal apical dendritic ROIs, the difference in responses to unexpected (*A-B-C-U-G*) and expected (*A-B-C-D-G*) sequences increased across days, reaching statistical significance by session 3. 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. 3A–B). 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 over time, though the effect was only statistically significant in L2/3 (Fig. 3A–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. 3A–C, top). These results indicate that the responses to the unexpected stimuli evolved differently from those 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, e.g., from [Deitch et al., 2020; 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. To test this possibility, we also examined the evolution of the responses of the same ROIs to a different stimulus, a visual flow stimulus, which, based on prior work, was unlikely to drive strong expectation violations [Zmarz and Keller, 2016]. We observed that the responses to this stimulus were much more stable over sessions (Fig. S5, S6A–B). This indicates that our observations of changes in the responses to the Gabor sequences were 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 to decrease for both regular sequence frames and unexpected frames (though these changes were not statistically significant, due at least in part to the relatively small number of tracked ROIs compared to the size of the full population; see Materials and Methods), whereas the responses to the unexpected frames increased in the distal apical dendrites (Fig. S4B).

Next, in order to understand how the responses to the expected *D* frames versus unexpected *U* frames evolved, we examined the USIs for the tracked ROIs over days. We found that in the somatic compartments, the USIs had converged towards zero over the three sessions (Fig. 3D–E, bottom). In contrast, in the distal apical dendritic compartments, the USIs had increased significantly over the three days. This effect was most prominent in L2/3-D (Fig. 3D–E, top). Notably, these effects were consistent across mice (Fig. S4C). 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. This discovery is consistent with the second and third predictions of the hierarchical learning hypothesis articulated in the Introduction, because we see changes over days and they are different in the distal apical dendrites and somata.

2.4 Responses to expected and unexpected stimuli predict subsequent response changes 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. 3), and that there is a difference in this evolution between ROIs receiving primarily bottom-up or top-down information (Fig. 3). These findings confirm three of the main predictions of our unexpected-event-driven sensory learning hypothesis. The final prediction we tested is that neural circuits use the cell-by-cell or distal apical dendritic segment-by-segment differences in responses

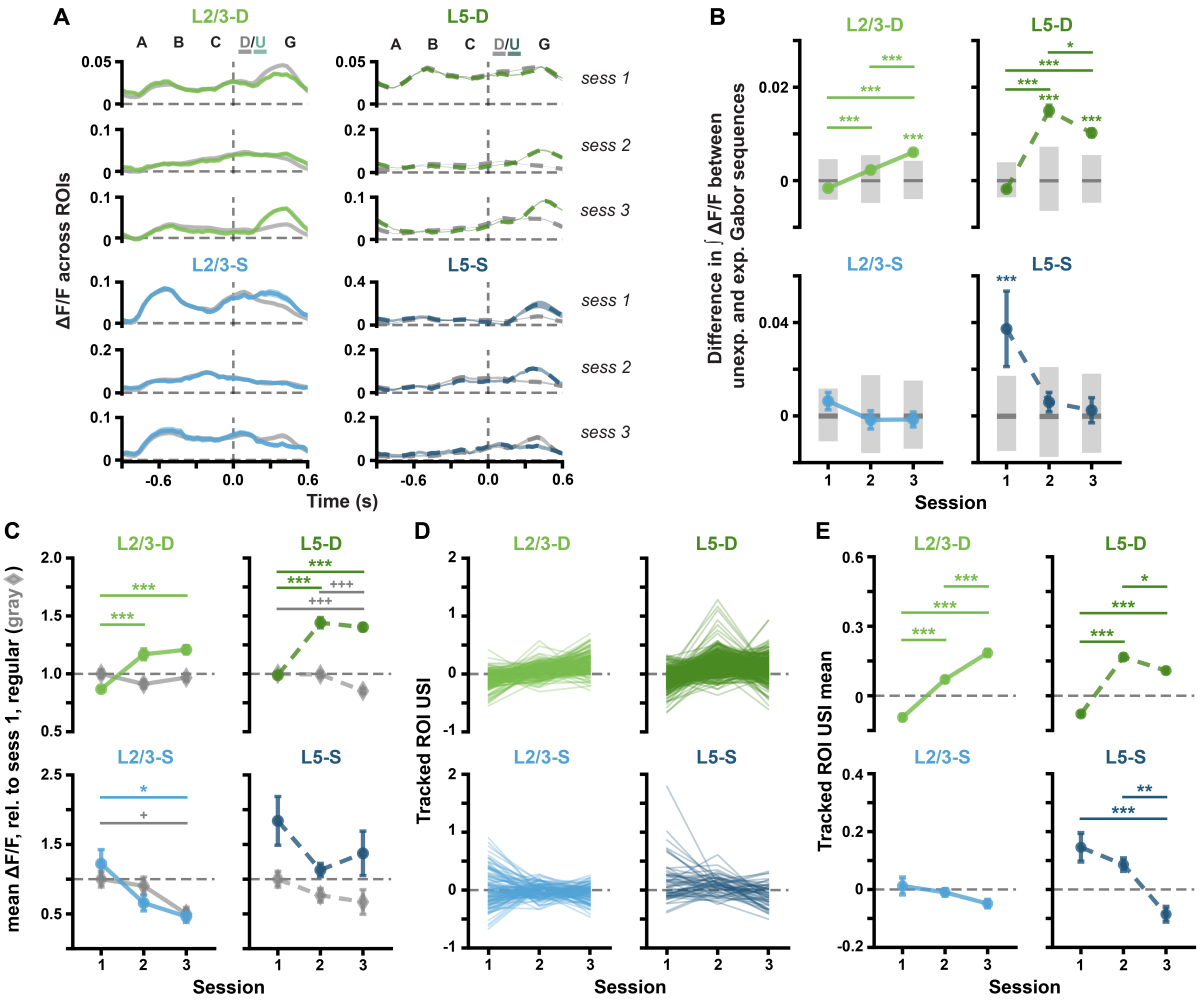

Figure 3: Unexpected Gabor sequences result in different $\Delta F/F$ and USI changes in different imaging planes.

(A) Mean (± SEM) $\Delta F/F$ response traces across ROI mean 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 (± SEM) differences across ROIs in the mean integrated response 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 (± 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*). Dashed horizontal lines mark the level of regular responses in session 1.
(D) Gabor sequence stimulus USIs for all tracked ROIs. Each line represents a single ROI's USIs over all 3 sessions.
(E) Gabor sequence stimulus USI means (± SEM) over tracked ROIs for data shown in (E).
*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ (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.

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 400 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 decreased over sessions until it was at or near chance level by session 3 (Fig. 4A, 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. 4A, top). Interestingly, these results parallel 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 fourth prediction held, i.e., whether the difference in responses to expected and unexpected stimuli could predict the subsequent changes in the neural responses over days. To achieve this goal, we analyzed the relationship between each ROI's USI in a given session and the subsequent change in that ROI's USI by the next session. Specifically, we examined the correlation between the USIs in each session and the changes in USIs between that session and the following one. Importantly, when one calculates a correlation like this, there is an inherent bias towards negative correlations in the analysis that results from the same variable being present in both values for which the correlation is calculated. To control for this, we used non–parametric tests to calculate the level of raw correlation for each ROI under the null hypothesis that the changes in USIs were not determined by the initial USIs. This allowed us to determine a normalized residual correlation for each ROI, calculated with respect to the level expected under the null hypothesis (Fig. 4B). Similar to our previous analyses, this was done by computing null distributions over raw correlations for each imaging plane through 10^4 random shuffles of the session labels (e.g., "session 1" or "session 2") associated with each USI value. From these null distributions, the magnitude and statistical significance of each imaging plane's normalized residual correlation was calculated.

As anticipated by the decoding analyses, which showed that information about the orientation of the unexpected stimulus was weaker by the second session for the somata, we found that in the somatic compartments there was a statistically significant negative normalized residual correlation between the USIs in session 1 and the change in USIs between sessions 1 and 2, but not between sessions 2 and 3 (Fig. 4C, bottom). This means that somatic ROIs with a stronger preference for the unexpected frames in session 1 had a greater tendency to decrease their selectivity for the unexpected frames between sessions 1 and 2. In contrast, in the distal apical dendritic ROIs, we found that there was a statistically significant positive normalized residual correlation between the USIs in session 2 and the change in USIs between sessions 2 and 3 (Fig. 4C, top). This means that dendritic ROIs with stronger preferences for the unexpected frames in session 2 had a greater tendency to increase their selectivity for the unexpected frames between sessions 2 and 3. Thus, the differences between the responses to the expected and unexpected events were predictive of how the selectivity of each ROI subsequently evolved over days.

As a control, we ran this same analysis on the visual flow stimulus, which was not expected to drive strong expectation violations [Zmarz and Keller, 2016], and had not induced clear changes in representations over sessions. As predicted, there were no statistically significant normalized residual correlations in response to this stimulus (Fig. S6F), demonstrating that our significant findings for the Gabor sequence stimulus were unlikely to be artefacts of our analysis or experimental methods. Altogether, these data provide support for the fourth prediction of hierarchical learning from unexpected events. Specifically, these results demonstrate that at a cell-by-cell or dendritic segment-by-segment level, the differences between responses to expected and unexpected stimuli can predict subsequent learning. This implicates the distinct responses to expected versus unexpected stimuli in the learning process, which is a key prediction of the hierarchical 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

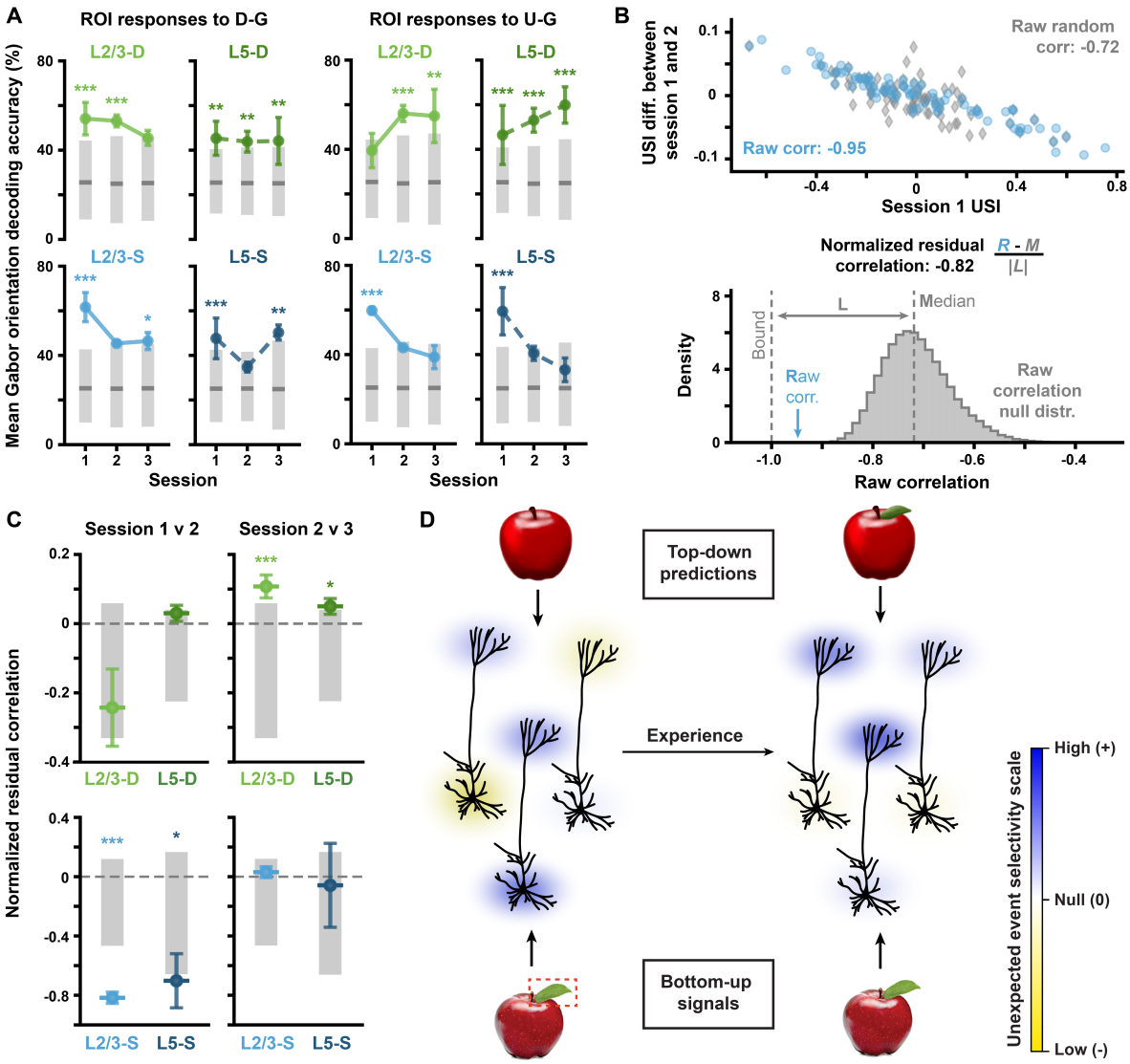

Figure 4: Unexpected Gabor sequences result in predictable $\Delta F/F$ changes in individual ROIs.

(A) Balanced accuracy (mean ± SEM over mice) for classifiers decoding mean Gabor patch orientations from ROI activity during *D-G* or *U-G* frames (2–3 mice per imaging plane, 400 random cross-validation splits per mouse, per session). Gray bars show median (dark) and adjusted 95% CIs (light), computed with shuffled orientation labels.
(B) Example illustration of how the normalized residual correlation between initial USIs and their subsequent change were computed from raw Pearson correlation coefficients. (Top) Scatter plots of true data (blue) and an example set with session labels shuffled (gray) for L2/3-S. (Bottom) Raw correlation and correlation distribution obtained by shuffling session labels. Residual correlations were computed as per the given formula (see also Equation 3), where we divided by $|L|$ to normalize the computed value such that it lay in [-1, 1].
(C) Normalized residual correlations (± bootstrapped SD) for each plane and session comparison, as defined in (B). Gray bars show adjusted 95% CIs.
(D) Illustration of a conceptual model based on our data of how unexpected events drive changes in the neural circuit. With experience, unexpected event selectivity in somata converges toward 0, whereas it increases in distal apical dendrites.
*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ (two-tailed, corrected).
See Table S1 for details of statistical tests and precise p-values for all comparisons.

a hierarchical model of the world by comparing predictions about sensory stimuli to the actual stimuli received from the world. This class of theories makes several predictions about neural responses and how they should evolve over time in response to expected versus unexpected stimuli. We tested four such predictions here, using chronic recordings in mouse VisP, and found strong evidence in support of each one. First, we observed that neurons responded systematically differently to expected versus unexpected stimuli, which is a precondition for learning from the 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 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 are shaped differently by the unsupervised learning process, which is a feature of learning in a hierarchical model. Finally, we found that the neurons displaying the largest sensitivity to the unexpected events also showed the largest subsequent changes in their response between imaging sessions. This final observation is the strongest test of the learning theories: namely, the sensitivity to unexpected events is predictive of subsequent changes during learning, on a cell-by-cell or dendritic segment-by-segment basis.

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

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

An open question in neuroscience is how the brain determines which cells and synapses need to be updated to improve hierarchical representations. This is known as the credit assignment problem. We argue that our data speak to this question. There are two broad classes of known algorithms for credit assignment in hierarchical neural networks. The first, and most common, involves calculating at each layer how the neural activity should change in order to improve the system's performance [Lillicrap et al., 2016, 2020]. Such "change-based" algorithms—of which backpropagation of error is an example—predict that neural signals for distinguishing unexpected from expected events will converge towards zero as a network learns about the unexpected events. This prediction arises because, once a network has learned about these events, no more changes are needed. In contrast, the second approach proposes that each layer should calculate what its activity should be, in order to improve the stimulus representation overall [Bengio, 2014; Bartunov et al., 2018]. These "target-based" algorithms—of which target propagation is an example—do not predict that the differences between expected and unexpected signals should shrink over time. Our data show that the somata, but not the distal apical dendrites, tend to show a reduction in the magnitude of the difference between the responses to expected and unexpected frames over time (Fig. 3). This observation suggests that the somatic compartments in the neocortical microcircuit might be better described by change-based credit assignment algorithms than target-based ones. Interestingly, this same statement does not apply to the distal apical dendrites: the differences between their responses to expected versus unexpected events do not converge towards zero over the multiple days of learning. The implication, therefore, is that somata and distal apical dendrites may be engaged in distinct calculations during learning. Interestingly, this differentiation has been a feature of a number of recent

theoretical models for biologically realistic deep learning [Guerguiev et al., 2017; Naud and Sprekeler, 2018; Sacramento et al., 2018; Federer and Zylberberg, 2018]. Our data show that these models are likely correct in supposing distinct roles for the distal apical dendrites in learning.

Thus, based on our data and previous results in the field, we propose a broad conceptual model illustrated in Fig. 4D. 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. 4D, 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. 4D, right). As a result, more accurate predictions are received at the distal apical dendrites, and so the somata show a decrease in selectivity to the previously unexpected events. In contrast, as the top-down predictions become better, the distal apical dendritic activity becomes more selective for these novel forms of stimuli.

Notably, our results do not support a simple version of predictive coding wherein neurons only communicate prediction errors. Although the unexpected event responses in the somata did decrease over time, the ones 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 with exposure to the unexpected stimuli goes counter to proposals implementing predictive coding by using the distal apical dendrite as a site for prediction error calculations [Sacramento et al., 2018; Whittington and Bogacz, 2019]. More experiments with simultaneous imaging of dendrites and cell bodies in different brain regions could help to clarify the distinct computational roles of these neuronal compartments.

There are a number of limitations to this work that must be recognized. First, we were not recording somata and distal apical dendrites in the same neurons. Thus, we cannot say with certainty that the differences between distal apical dendritic and somatic responses hold within individual cells. Indeed, there is data to suggest that coupling between the distal apical dendrites and the somata can sometimes be strong. Nonetheless, we saw clear differences in the evolution of selectivity for unexpected events over time between the somatic and distal apical dendritic compartments. These observations were consistent across mice (Fig. S4C) and it thus seems unlikely that these results would not hold within individual neurons. At the same time, future work using simultaneous multi-plane imaging could help answer these questions.

Second, given the nature of our visual stimuli we were unable to measure either the classical receptive fields or the orientation tuning of the neurons. As such, we cannot state with certainty whether these factors could explain the differences in how individual cells responded to expected and unexpected stimuli. However, we observed our results in aggregate across large populations of recorded neurons, presumably with diverse orientation tuning properties and receptive fields. Thus, it is unlikely that idiosyncracies of individual neurons' orientation selectivities could account for the unexpected event responses. This assertion is supported by our finding that even when we only compared responses for expected and unexpected frames with the same mean orientation we still observed significant differences in the responses. Moreover, we observed significant changes in unexpected event selectivity over days, whereas classic receptive fields and orientation tuning of neurons in mouse V1 are known to be relatively stable over these timescales [Montijn et al., 2016].

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

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

Fifth, though we examined the distal apical dendrites separately from the somata in order to examine potential differences in the processing of top-down and bottom-up inputs, an ideal experiment would record simultaneously from other higher-order brain regions and their projections into visual cortex. This would help determine whether the signals we saw in the distal apical dendrites were being calculated locally or in other regions. These experiments would be technically challenging, but potentially feasible in the not-too-distant

future given recent technical advances in multi-plane mesoscope imaging.

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

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

4 Materials and Methods

4.1 Experimental animals and calcium imaging

The data used in this paper were 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 at 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: $n = 3$). Amongst these, only one L2/3-S mouse and one L5-S mouse had incomplete series, with sessions 2–3 and session 3 missing, respectively, due to not passing quality control. 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], 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. 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, 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,[1] and the full dataset is publicly available in Neurodata Without Borders (NWB) format [Ruebel et al., 2019] in the DANDI Archive.[2]

4.2 Visual stimulation

During each habituation and imaging session, mice viewed the Gabor sequence stimulus, as well as an additional stimulus, which was a visual flow stimulus. The stimuli were presented consecutively for an equal amount of time and in random order. They appeared on a grayscreen background and were projected on a flat 24-inch monitor positioned 10 cm from the right eye. The monitor was rotated and tilted to appear perpendicular to the optic axis of the eye, and the stimuli were warped spatially to mimic a spherical projection screen. Whereas habituation sessions increased in duration over days from 10 to 60 minutes, optical imaging sessions always lasted 70 minutes, comprising 34 minutes of Gabor sequence stimulus and 17 minutes of visual flow stimulus in each direction. Each stimulus period was flanked by one or 30 seconds of grayscreen for the habituation and optical imaging sessions, respectively.

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

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

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

The stimuli were generated using Python 2.7 [Van Rossum and Drake Jr, 1995] custom scripts based on PsychoPy 1.82.01 [Peirce, 2009] and CamStim 0.2.4, which was developed and shared by the Allen Institute for

[1] https://github.com/jeromelecoq/allen_openscope_metadata/tree/master/projects/credit_assignement
[2] https://gui.dandiarchive.org/#/dandiset/000037

Brain Science. Code, instructions to reproduce the stimuli, and example videos are available on Github.[3]

4.3 Statistical analyses

For most analyses, mean ± standard error of the mean (SEM) is reported. In cases where the error could not be directly measured over the sample, e.g., the percentage of significant ROI USIs reported in Fig. 2F, a bootstrapped estimate of the error was obtained by resampling the data with replacement 10^4 times. In these cases, the standard deviation (SD) over the bootstrapped sample is plotted instead, and this is visually signaled by the use of broader error caps (Fig. 2F–G, 4C, S6B, E, F).

Significance tests, unless otherwise indicated, were computed non–parametrically using permutation tests with 10^5 shuffles to construct null distributions, based on which confidence intervals (CIs) could be estimated. Where p-values are reported, they are two-tailed, and Bonferroni-corrected for multiple comparisons to reduce the risk of Type I errors (false positives). Where 95% CIs are plotted, they are equivalently adjusted using a Bonferroni correction. An exception was made for Fig. 2I, which reports the relationship between the stimuli and behavioral data. Here, Type II errors (false negatives) were considered of greater concern, and thus we reported raw two-tailed p-values in the panel itself. Details of the statistical analyses for all figures, including number of comparisons and corrected p-values, are presented in Table S1.

4.4 Running and pupil analysis

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

To track pupil diameter during imaging sessions, an infrared LED illuminated the eye ipsilateral to the monitor (right eye), allowing infrared videos to be recorded (Fig. 2H, Supp. Video 3) [Allen Institute for Brain Science, 2017]. We trained a DeepLabCut model from ∼200 manually labeled examples to automatically label points around the eye, from which we estimated the pupil diameter (∼0.01 mm per pixel conversion) [Mathis et al., 2018].[4] We omitted from analyses outlier frames, defined as resulting from a single-frame diameter change of at least 0.05 mm, which usually resulted from blinking.

Each datapoint in Fig. 2I corresponds to the difference in the mean running velocity or pupil diameter for one block between the unexpected and preceding expected Gabor sequences during session 1, with all blocks being pooled across mice. We computed p-values by comparing the mean difference over all blocks for each plane to a distribution of mean differences, obtained by shuffling the expected and unexpected labels 10^4 times and calculating the mean difference over all blocks each time.

4.5 ROI tracking across sessions

To track ROIs across days, we employed a custom-modified version of the ROI-matching package developed to track cell bodies across multiple recording days by the Allen Institute for Brain Science [de Vries et al., 2020]. This pipeline implements the enhanced correlation coefficient image registration algorithm to align ROI masks and the graph-theoretic blossom algorithm to optimize the separation and degree of overlap between pairwise matches, as well as the number of matches across all provided sessions [Evangelidis and Psarakis, 2008]. This process produced highly plausible matches for the somatic ROIs; however, it provided some implausible matches for the smaller and more irregularly shaped dendritic ROIs. We therefore further constrained the putative matches to those that overlapped by at least 30% (no such constraint was built in), and merged results across all session orderings (e.g., 1-2-3, 1-3-2, 3-1-2), eliminating any conflicting matches, i.e., non–identical matchings that shared ROIs. In total, the modified matching algorithm produced ∼50–300 highly plausible matched ROIs per plane, i.e., ∼22–33% of the theoretical maximum number of trackable ROIs (L2/3-D: $n = 166$, L2/3-S: $n = 113$, L5-D: $n = 346$, L5-S: $n = 56$) (Fig. 1E, S1, S7).

[3] https://github.com/colleenjg/cred_assign_stimuli

[4] https://allensdk.readthedocs.io/en/latest/allensdk.internal.brain_observatory.eye_calibration.html

4.6 Fluorescence trace analysis

For all results except those presented in Fig. 4A, S4A, C, and S6D, ROIs were pooled across all mice within an imaging plane for analyses. To enable ROI pooling across mice within imaging planes, each ROI's $\Delta F/F$ trace was scaled using robust standardization, i.e., by subtracting the median and then dividing by the interpercentile range between the 5th and 95th percentiles.

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

$$\text{USI} = \frac{\mu_{\text{unexpected}} - \mu_{\text{expected}}}{\sqrt{\frac{1}{2}\left(\sigma^2_{\text{expected}} + \sigma^2_{\text{unexpected}}\right)}}, \tag{1}$$

where the means (μ_{expected} and $\mu_{\text{unexpected}}$) and variances ($\sigma^2_{\text{expected}}$ and $\sigma^2_{\text{unexpected}}$) were calculated across integrated $\Delta F/F$ responses to the expected and unexpected events, respectively. For the Gabor sequences, expected events responses were defined as those spanning *D*-*G* frames, and unexpected events were defined as those spanning *U*-*G* frames, with each event therefore spanning 600 ms. Indeed, *G* frames were included in these events, as they did not introduce any new stimuli, but did consistently show persisting ROI responses to *D* or *U* frames (Fig. 2C). For the visual flow stimulus, expected events were defined as the last 2 seconds of expected flow before unexpected flow onset (at which point 25% of the squares reversed direction), while unexpected events were defined as the first 2 seconds of unexpected flow (Fig. S5B). For each ROI, in addition to the true USI, a null distribution over USIs was obtained by randomly reassigning the expected and unexpected event labels to each response 10^4 times. USIs were deemed significantly low if they lay below the 2.5^{th} percentile, and significantly high if they lay above the 97.5^{th} percentile of their null distribution (Fig. 2D).

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

For Fig. 3A-B, ROI responses and differences in responses to full expected (*A*-*B*-*C*-*D*-*G*) and unexpected (*A*-*B*-*C*-*U*-*G*) sequences were obtained by first taking the mean $\Delta F/F$ for each ROI across Gabor sequences. Mean $\Delta F/F$ ± SEM traces were then computed across ROIs and plotted for each session and imaging plane (Fig. 3A). For Fig. 3B, the differences in the traces plotted in (Fig. 3A) were quantified by integrating the mean $\Delta F/F$ responses over time for each ROI. Mean differences ± SEM between expected and unexpected sequence responses were then calculated across ROIs and plotted for each session and imaging plane. To further compare ROI responses to regular (*A*-*B*-*C*) and unexpected (*U*-*G*) stimuli, for each ROI, a mean $\Delta F/F$ was calculated within each Gabor sequence, and then across sequences (Fig. 3C). The mean $\Delta F/F$ values thus obtained for each ROI over a given session were then normalized by dividing by the mean $\Delta F/F$ for regular stimuli across all ROIs from the same mouse in session 1. These normalized means ± SEM over ROIs were then plotted for each session and plane. Absolute fractional differences between sessions in the responses to unexpected stimuli (Fig. S6B) or in USIs (Fig. S6E) were defined as

$$\left|\frac{\mu_3 - \mu_1}{\mu_1}\right|, \tag{2}$$

where the subscripts indicate the session over which the mean μ is computed. For Fig. S6B, μ is the mean of the $\Delta F/F$ values over all ROIs for the given plane or pooled over all planes, as indicated, for unexpected sequences. As in Fig. 3C, the $\Delta F/F$ values were calculated relative to the mean expected $\Delta F/F$ values on session 1 for each mouse. For Fig. S6E, μ is the mean of the absolute values of the USIs for the given plane or pooled over all planes for unexpected sequences. Significance tests comparing both Gabor sequence and visual flow stimulus results were assessed by permuting stimulus labels to compute adjusted 95% CIs over stimulus differences expected by chance (Fig. 3C, E, S6A, B, E).

For the orientation decoding analyses, linear logistic regressions were trained with an L_2 penalty on the multinomial task of classifying the mean Gabor patch orientation for *D*-*G* frames {0°, 45°, 90°, 135°} or *U*-*G* frames {90°, 135°, 180°, 225°}. Balanced classifier accuracy was evaluated on the test sets of 400 random cross-validation 75:25 train:test splits of the dataset for each mouse. Input data consisted of the $\Delta F/F$ responses for all ROIs together across *D*-*G* or *U*-*G* frames (600 ms). The traces were standardized as described above, but using statistics drawn from the training data only. Mean balanced accuracy across dataset splits was calculated

for each mouse, and the mean (± SEM) balanced accuracy across mice was plotted for each session and plane. To estimate chance accuracy, shuffled classifier performances were evaluated on 10^5 random cross-validation dataset splits for each mouse. These classifiers were trained as above, but for each split, the training set orientation targets were shuffled randomly. Null distributions over mean performance were obtained by averaging classifier accuracy for each split across mice, from which adjusted 95% CIs over accuracy levels expected by chance were calculated for each session and plane (Fig. 4A).

Normalized residual correlations (Fig. 4C, S6F) were calculated to measure the correlation between the ROI USIs in an imaging plane in one session i (USI_i) and the change in the USIs by the next session $i+1$ ($\text{USI}_{i+1} - \text{USI}_i$). However, since the variables regressed were of the form $y - x$ and x, strong negative correlations could reflect their algebraic relationship alone. To compensate for this bias, we computed a null distribution over raw correlations for each plane, obtained by permuting the session labels 10^5 times and recomputing the raw correlation each time. The true raw correlation was then compared to the distribution obtained under the null hypothesis that there was no session-to-session structure in the USIs and their differences. For each plane, a normalized residual correlation (N) was calculated using Equation 3:

$$N = \frac{R - M}{|L|}, \tag{3}$$

where R is the true raw correlation, M is the median of the null distribution, and $L = B - M$, i.e., the distance between M and the nearest theoretical correlation bound B (-1 or +1). Normalizing by $|L|$ ensured that the normalized residual correlations lay in [-1, 1]. Finally, bootstrapped SDs over correlations for each plane were estimated, and adjusted 95% CIs were computed by taking the appropriate null distribution percentile values and substituting them for R in Equation 3.

5 Analysis software

Analyses were performed in Python 3.6 [Van Rossum and Drake, 2009] with custom scripts that are freely available on GitHub,[5] and were developed using the following packages: NumPy [Harris et al., 2020], SciPy [Jones et al., 2001], Pandas [McKinney et al., 2010], Matplotlib [Hunter, 2007], Scikit-learn 0.21.1 [Pedregosa et al., 2011], and the AllenSDK 1.6.0.[6] Dendritic segmentation was run in Matlab 2019a [MATLAB, 2019] using the robust estimation algorithm developed by [Inan et al., 2017]. Pupil tracking was performed using DeepLabCut 2.0.5 [Mathis et al., 2018]. ROIs were matched across sessions using a custom-modified version of the n-way cell matching package developed by the Allen Institute.[7]

6 Acknowledgements

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

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

This work was enabled by the resources provided by Compute Ontario and Compute Canada.[8]

[5] https://github.com/colleenjg/OpenScope_CA_Analysis
[6] https://github.com/AllenInstitute/AllenSDK
[7] https://github.com/AllenInstitute/ophys_nway_matching/tree/main/nway
[8] www.computeontario.ca and www.computecanada.ca

7 Funding

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

References

Allen Institute for Brain Science (2017). Visual coding overview. Technical report, http://observatory.brain-map.org/visualcoding/.

Bartunov, S., Santoro, A., Richards, B., Marris, L., Hinton, G. E., and Lillicrap, T. (2018). Assessing the scalability of biologically-motivated deep learning algorithms and architectures. In *Advances in Neural Information Processing Systems*, pages 9368–9378.

Beaulieu, C. and Cynader, M. (1990). Effect of the richness of the environment on neurons in cat visual cortex. I. Receptive field properties. *Developmental Brain Research*, 53(1):71–81.

Bengio, Y. (2014). How auto-encoders could provide credit assignment in deep networks via target propagation. *arXiv preprint arXiv:1407.7906*.

Budd, J. M. (1998). Extrastriate feedback to primary visual cortex in primates: a quantitative analysis of connectivity. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1400):1037–1044.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*.

Christensen, E. and Zylberberg, J. (2020). Models of the primate ventral stream that categorize and visualize images. *bioRxiv:2020.02.21.958488*.

Dayan, P., Hinton, G. E., Neal, R. M., and Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, 7(5):889–904.

de Vries, S. E., Lecoq, J. A., Buice, M. A., Groblewski, P. A., Ocker, G. K., Oliver, M., Feng, D., Cain, N., Ledochowitsch, P., Millman, D., et al. (2020). A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature Neuroscience*, 23(1):138–151.

Deitch, D., Rubin, A., and Ziv, Y. (2020). Representational drift in the mouse visual cortex. *bioRxiv:2020.10.05.327049*.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Evangelidis, G. D. and Psarakis, E. Z. (2008). Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1858–1865.

Federer, C. and Zylberberg, J. (2018). A self-organizing short-term dynamical memory network. *Neural Networks*, 106:30–41.

Friston, K. and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521):1211–1221.

Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clinical Neurophysiology*, 120(3):453–463.

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., et al. (2020). Bootstrap your own latent - a new approach to self-supervised learning. *Advances in Neural Information Processing Systems*, 33.

Guerguiev, J., Lillicrap, T. P., and Richards, B. A. (2017). Towards deep learning with segregated dendrites. *ELife*, 6:e22901.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., et al. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

Hawkins, J. and Blakeslee, S. (2004). *On intelligence*. Macmillan.

Higgins, I., Stringer, S., and Schnupp, J. (2017). Unsupervised learning of temporal features for word categorization in a spiking neural network model of the auditory brain. *PLoS One*, 12(8):e0180174.

Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507.

Homann, J., Koay, S. A., Glidden, A. M., Tank, D. W., and Berry, M. J. (2017). Predictive coding of novel versus familiar stimuli in the primary visual cortex. *bioRxiv:197608*.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

Hyvärinen, A., Sasaki, H., and Turner, R. (2019). Nonlinear ICA using auxiliary variables and generalized contrastive learning. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 859–868.

Inan, H., Erdogdu, M. A., and Schnitzer, M. (2017). Robust estimation of neural signals in calcium imaging. In *Advances in Neural Information Processing Systems*, pages 2901–2910.

Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open source scientific tools for Python.

Jordan, R. and Keller, G. B. (2020). Opposing influence of top-down and bottom-up input on excitatory layer 2/3 neurons in mouse primary visual cortex. *Neuron*.

Keller, G. B., Bonhoeffer, T., and Hübener, M. (2012). Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron*, 74(5):809–815.

Konkle, T. and Alvarez, G. (2020). Deepnets do not need category supervision to predict visual system responses to objects. *Journal of Vision*, 20(11):498–498.

Kumaran, D. and Maguire, E. A. (2006). An unexpected sequence of events: mismatch detection in the human hippocampus. *PLoS Biology*, 4(12):e424.

Larkum, M. E. (2013a). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends in Neurosciences*, 36(3):141–151.

Larkum, M. E. (2013b). The yin and yang of cortical layer 1. *Nature Neuroscience*, 16(2):114–115.

Larochelle, H. and Hinton, G. E. (2010). Learning to combine foveal glimpses with a third-order Boltzmann machine. In *Advances in Neural Information Processing Systems*, pages 1243–1251.

Lee, T. S. and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, 20(7):1434–1448.

Lillicrap, T. P., Cownden, D., Tweed, D. B., and Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*, 7(1):1–10.

Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., and Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, pages 1–12.

Lotter, W., Kreiman, G., and Cox, D. (2016). Deep predictive coding networks for video prediction and unsupervised learning. *arXiv preprint arXiv:1605.08104*.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., and Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9):1281–1289.

MATLAB (2019). *9.6.0.2030181 (R2019a)*. The MathWorks Inc., Natick, MA.

McKinney, W. et al. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX.

Millman, D. J., Ocker, G. K., Caldejon, S., Larkin, J. D., Lee, E. K., Luviano, J., Nayan, C., Nguyen, T. V., North, K., Seid, S., et al. (2020). VIP interneurons in mouse primary visual cortex selectively enhance responses to weak but specific stimuli. *Elife*, 9:e55130.

Montijn, J. S., Meijer, G. T., Lansink, C. S., and Pennartz, C. M. (2016). Population-level neural codes are robust to single-neuron variability from a multidimensional coding perspective. *Cell Reports*, 16(9):2486–2498.

Naud, R. and Sprekeler, H. (2018). Sparse bursts optimize information transmission in a multiplexed neural code. *Proceedings of the National Academy of Sciences*, 115(27):E6329–E6338.

Niell, C. M. and Stryker, M. P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, 65(4):472–479.

Orlova, N., Tsyboulski, D., Najafi, F., Seid, S., Kivikas, S., Griffin, F., Leon, A., L'Heureux, Q., North, K., Swapp, J., et al. (2020). Multiplane mesoscope reveals distinct cortical interactions following expectation violations. *bioRxiv:2020.10.06.328294*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2:10.

Press, C., Kok, P., and Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences*, 24(1):13–24.

Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87.

Ruebel, O., Tritt, A., Dichter, B., Braun, T., Cain, N., Clack, N., Davidson, T. J., Dougherty, M., Fillion-Robin, J.-C., Graddis, N., et al. (2019). NWB:N 2.0: An accessible data standard for neurophysiology. *bioRxiv:523035*.

Rule, M. E., O'Leary, T., and Harvey, C. D. (2019). Causes and consequences of representational drift. *Current Opinion in Neurobiology*, 58:141–147.

Sacramento, J., Ponte Costa, R., Bengio, Y., and Senn, W. (2018). Dendritic cortical microcircuits approximate the backpropagation algorithm. *Advances in Neural Information Processing Systems*, 31:8721–8732.

Salkoff, D. B., Zagha, E., McCarthy, E., and McCormick, D. A. (2020). Movement and performance explain widespread cortical activity in a visual detection task. *Cerebral Cortex*, 30(1):421–437.

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, 112:92–97.

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., and Harris, K. D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364(6437).

van den Oord, A., Li, Y., and Vinyals, O. (2018). Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Van Rossum, G. and Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace, Scotts Valley, CA.

Van Rossum, G. and Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Wayne, G., Hung, C.-C., Amos, D., Mirza, M., Ahuja, A., Grabska-Barwinska, A., Rae, J., Mirowski, P., Leibo, J. Z., Santoro, A., et al. (2018). Unsupervised predictive memory in a goal-directed agent. *arXiv preprint arXiv:1803.10760*.

Whittington, J. C. and Bogacz, R. (2017). An approximation of the error backpropagation algorithm in a predictive coding network with local Hebbian synaptic plasticity. *Neural Computation*, 29(5):1229–1262.

Whittington, J. C. and Bogacz, R. (2019). Theories of error back-propagation in the brain. *Trends in Cognitive Sciences*, 23(3):235–250.

Woloszyn, L. and Sheinberg, D. L. (2012). Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. *Neuron*, 74(1):193–205.

Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M., DiCarlo, J., and Yamins, D. (2020). Unsupervised neural network models of the ventral visual stream. *bioRxiv:2020.06.16.155556*.

Zmarz, P. and Keller, G. B. (2016). Mismatch receptive fields in mouse visual cortex. *Neuron*, 92(4):766–772.

Zylberberg, J., Murphy, J. T., and DeWeese, M. R. (2011). A sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of V1 simple cell receptive fields. *PLoS Computational Biology*, 7(10):e1002250.

8 Supplemental Info

8.1 Supplemental videos

Supp. Video 1: Sample two-photon recordings for each imaging plane
Supp. Video 2: Sample of a running recording
Supp. Video 3: Sample of an annotated pupil recording
Supp. Video 4: Gabor sequence stimulus example
Supp. Video 5: Visual flow stimulus example

8.2 Supplemental figures

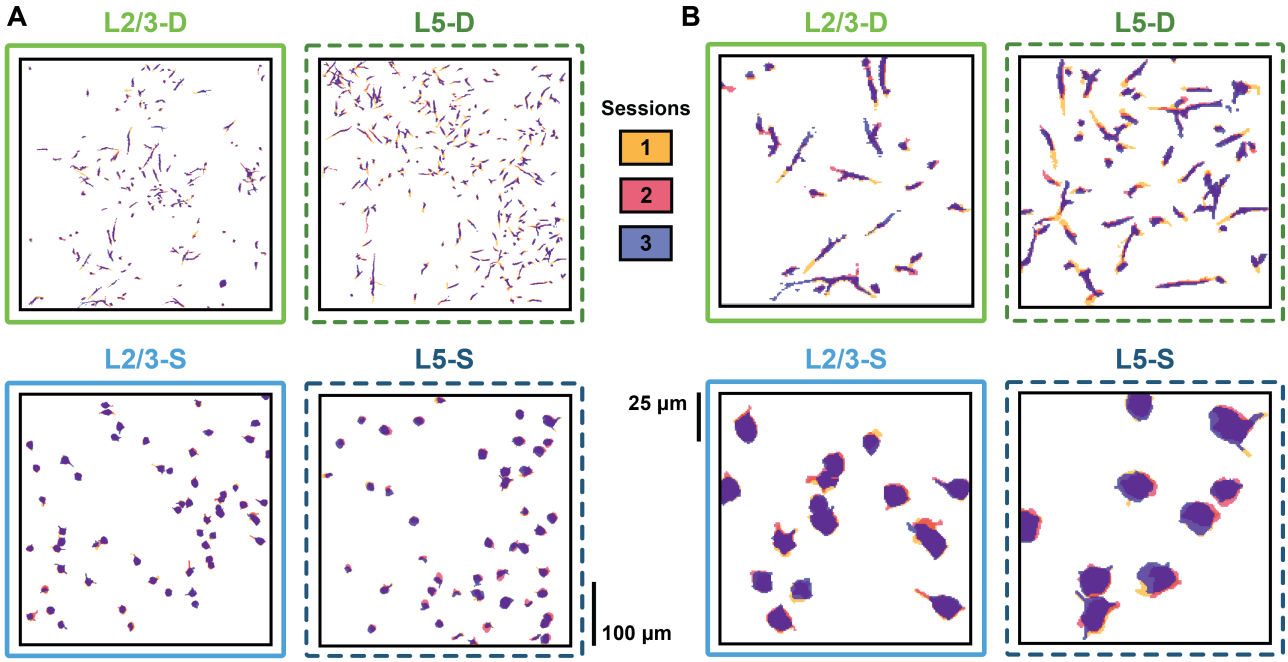

Figure S1: ROIs are successfully tracked in each plane.

(A) Full field of view overlays of ROIs tracked across all 3 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 produces highly plausible ROI matches across all three sessions in each imaging plane.

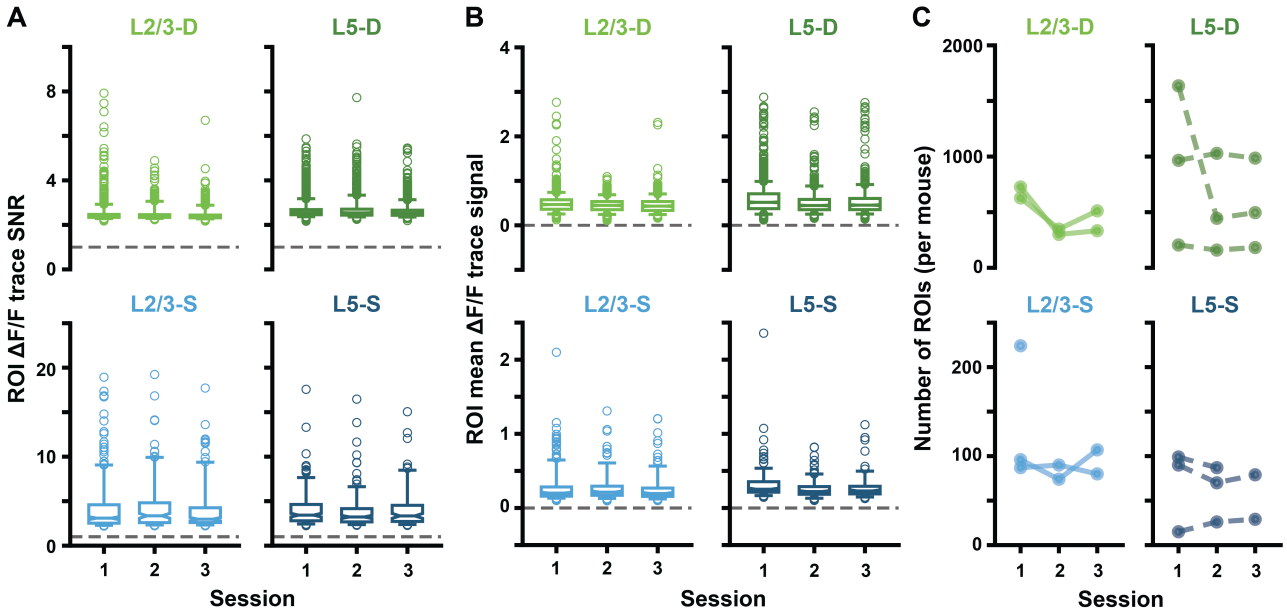

Figure S2: ROI SNR, signal and numbers are consistent across sessions.

(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 were generally stable across sessions for each mouse.

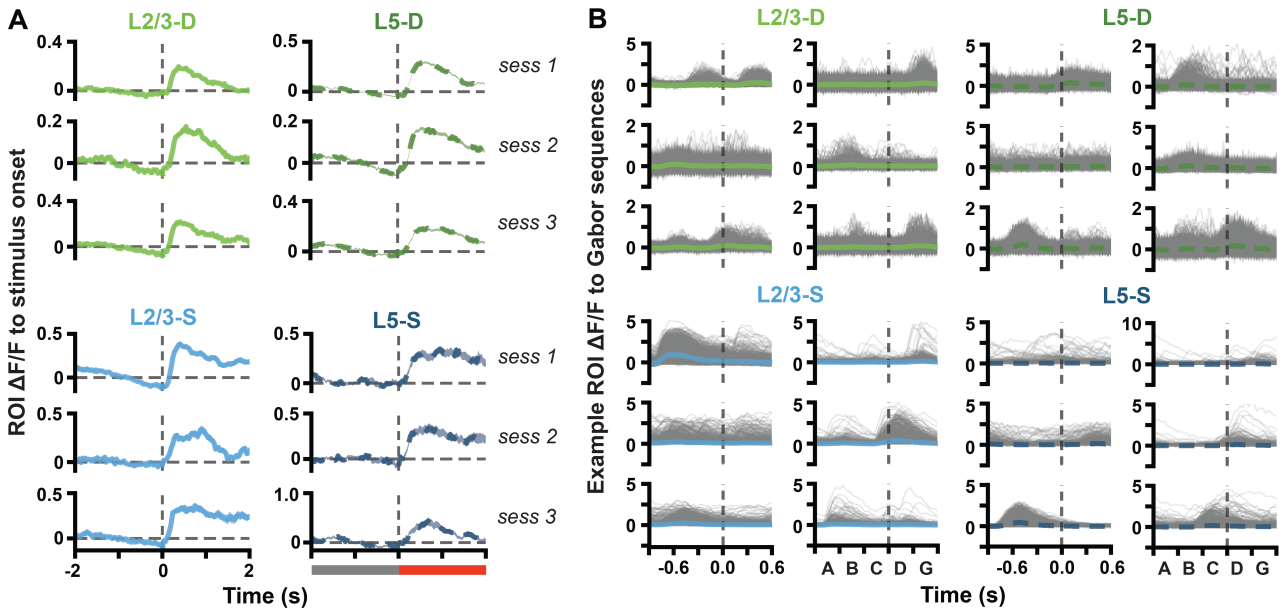

Figure S3: ROIs are responsive to stimulus onset, and to Gabor sequences.

(A) Mean (± 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 (± SEM) $\Delta F/F$ responses across sequences are plotted in blue or green. Dashed vertical lines mark onset of D/U frames. Plotted ROIs were randomly selected from ROIs deemed consistently responsive to Gabor sequences, based on the following criteria: (1) the mean cross-correlation across sequence responses, (2) the standard deviation of the mean response, and (3) the skew of the mean response were each above the 75th percentile for ROIs whose SNR was above the session median.
In each imaging plane, numerous ROIs were found which were responsive to various components of the Gabor sequences.

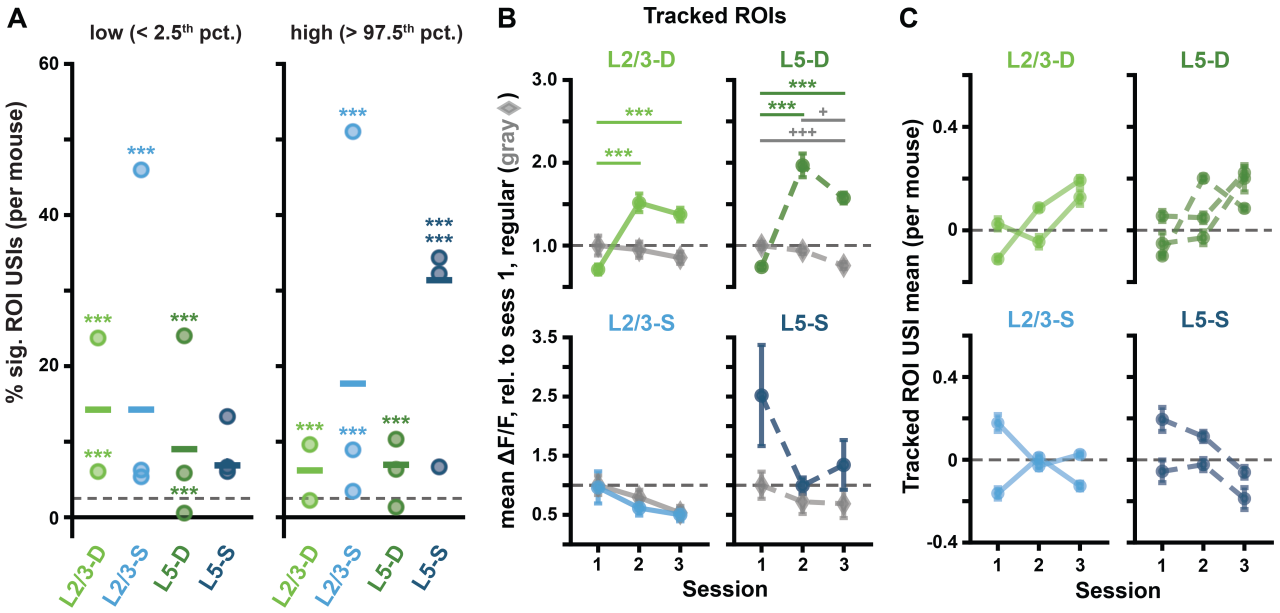

Figure S4: ROI responses to unexpected Gabor sequences are consistent in tracked ROIs and across mice.

(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 plotted in Fig. 2F. Dashed horizontal lines mark the theoretical chance level (2.5%).
Results are consistent with those pooled across mice, with 9 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 (± 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. 3C. Dashed horizontal lines mark the level of regular responses in session 1.
Results are consistent with the full ROI population results (Fig. 3C).
(C) Gabor sequence stimulus USI means (± SEM) over tracked ROIs as in Fig. 3E, but split by mouse.
Results are consistent with those pooled across mice (Fig. 3E).
*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ (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.

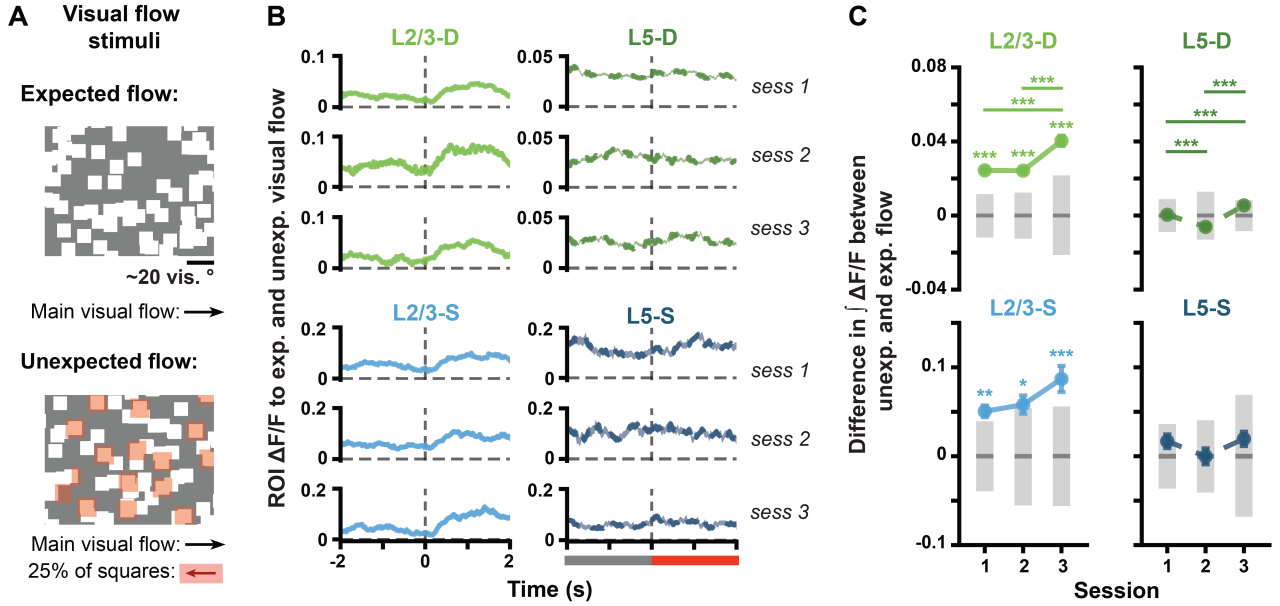

Figure S5: L2/3 ROIs respond to the onset of unexpected visual flow.

(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 and Methods, and Supp. Video 5).
(B) Mean (± SEM) $\Delta F/F$ response traces across ROI mean 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 (± SEM) differences across ROIs in the mean integrated response 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 visuomotor 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.

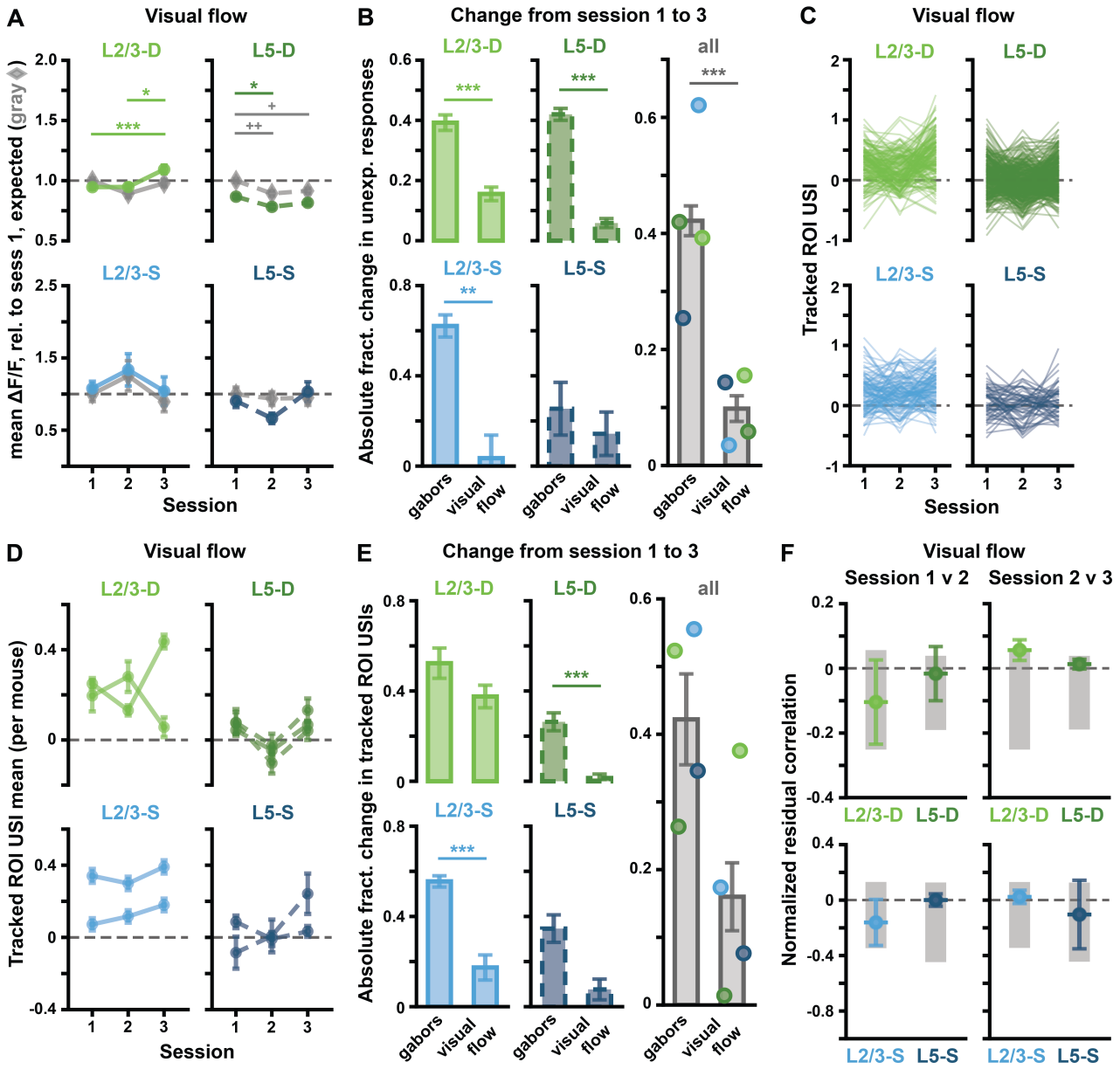

Figure S6: Unexpected visual flow sequences do not result in the same predictable $\Delta F/F$ changes across sessions as unexpected Gabor sequences do.

(A) Mean (± 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. Dashed horizontal lines mark the level of regular responses in session 1. Corresponds to Fig. 3C for Gabor sequences.
(B) Absolute fractional change (± 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 3 sessions. Corresponds to Fig. 3D for Gabor sequences.
(D) Visual flow stimulus USI means (± SEM) over tracked ROIs, split by mouse. Corresponds to Fig. S4C for Gabor sequences.

(Fig. S6 caption, cont'd)
(E) Similar to (B), but here the mean (± bootstrapped SD) absolute fractional changes in USIs from session 1 to 3 across tracked ROIs are plotted (see Equation 2).
In L5-D and L2/3-S, 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) Normalized residual correlations (± bootstrapped SD) for each plane and session comparison, as defined in Fig. 4B and Equation 3. Gray bars show adjusted 95% CIs. Corresponds to Fig. 4C for Gabor sequences.
Unlike the Gabor stimulus, for the visual flow stimulus, no significant correlations are observed between USIs and subsequent changes in USIs from one session to another.
*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ (two-tailed, corrected).
$^{+}$: $p < 0.05$, $^{++}$: $p < 0.01$, $^{+++}$: $p < 0.001$ (two-tailed, corrected), for expected stimulus comparisons (gray) in (A).
See Table S1 for details of statistical tests and precise p-values for all comparisons.

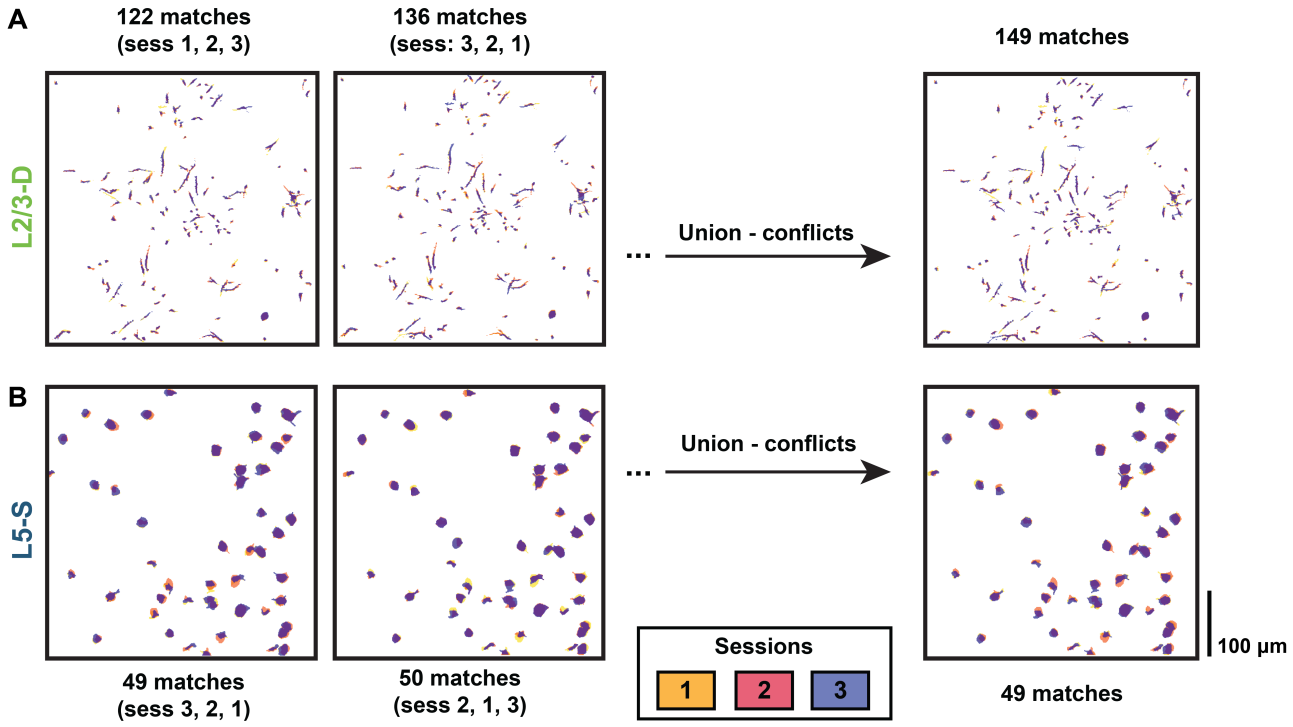

Figure S7: Dendritic ROI matches vary more with session ordering than somatic ROI matches do.

(A) Example L2/3-S mouse with ROIs matched across sessions. The order in which the session images are aligned affects which ROIs are matched. (Left) Permutation with the smallest number of matched ROIs. (Middle) Permutation with the largest number of matched ROIs. (Right) Combining 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.
(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 enable one of the pairwise matches to be identified as a conflict and removed.

8.3 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 left	L2/3-D to null		binomial null CI	8 comparisons	< 0.001	$p < 0.001$
		L2/3-S to null				< 0.001	$p < 0.001$
		L5-D to null				< 0.001	$p < 0.001$
		L5-S to null				0.011	$p < 0.05$
	F right	L2/3-D to null				< 0.001	$p < 0.001$
		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 left	L2/3-D to null		binomial null CI	8 comparisons	< 0.001	$p < 0.001$
		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 right	L2/3-D to null				< 0.001	$p < 0.001$
		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	$p < 0.001$
	I left	L2/3-D to null		10^4 permutations	4 comparisons	1.000	n.s.
		L2/3-S to null				1.000	n.s.
		L5-D to null				1.000	n.s.
		L5-S to null				1.000	n.s.
	I right	L2/3-D to null		10^4 permutations	4 comparisons	1.000	n.s.
		L2/3-S to null				0.802	n.s.
		L5-D to null				1.000	n.s.
		L5-S to null				1.000	n.s.
3	B	L2/3-D	sess. 1 to null	10^5 permutations	24 comparisons	1.000	n.s.
			sess. 2 to null			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			1.000	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			1.000	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			0.019	$p < 0.05$
		L5-S	sess. 1 to null			< 0.001	$p < 0.001$
			sess. 2 to null			1.000	n.s.
			sess. 3 to null			1.000	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.
(3)	(B)	(L5-S)	sess. 1 vs. 2		(10^5	(24	0.755	n.s.
			sess. 1 vs. 3		permutations)	comparisons)	0.974	n.s.
			sess. 2 vs. 3				1.000	n.s.
	C	L2/3-D	sess. 1 vs. 2	reg.	10^5	24	1.000	n.s.
				unexp.	permutations	comparisons	< 0.001	$p < 0.001$
			sess. 1 vs. 3	reg.			1.000	n.s.
				unexp.			< 0.001	$p < 0.001$
			sess. 2 vs. 3	reg.			1.000	n.s.
				unexp.			1.000	n.s.
		L2/3-S	sess. 1 vs. 2	reg.			1.000	n.s.
				unexp.			0.558	n.s.
			sess. 1 vs. 3	reg.			0.034	$p < 0.05$
				unexp.			0.012	$p < 0.05$
			sess. 2 vs. 3	reg.			1.000	n.s.
				unexp.			1.000	n.s.
		L5-D	sess. 1 vs. 2	reg.			1.000	n.s.
				unexp.			< 0.001	$p < 0.001$
			sess. 1 vs. 3	reg.			< 0.001	$p < 0.001$
				unexp.			< 0.001	$p < 0.001$
			sess. 2 vs. 3	reg.			< 0.001	$p < 0.001$
				unexp.			1.000	n.s.
		L5-S	sess. 1 vs. 2	reg.			1.000	n.s.
				unexp.			1.000	n.s.
			sess. 1 vs. 3	reg.			1.000	n.s.
				unexp.			1.000	n.s.
			sess. 2 vs. 3	reg.			1.000	n.s.
				unexp.			1.000	n.s.
	E	L2/3-D	sess. 1 vs. 2		10^5	12	< 0.001	$p < 0.001$
			sess. 1 vs. 3		permutations	comparisons	< 0.001	$p < 0.001$
			sess. 2 vs. 3				< 0.001	$p < 0.001$
		L2/3-S	sess. 1 vs. 2				1.000	n.s.
			sess. 1 vs. 3				0.319	n.s.
			sess. 2 vs. 3				1.000	n.s.
		L5-D	sess. 1 vs. 2				< 0.001	$p < 0.001$
			sess. 1 vs. 3				< 0.001	$p < 0.001$
			sess. 2 vs. 3				0.013	$p < 0.05$
		L5-S	sess. 1 vs. 2				1.000	n.s.
			sess. 1 vs. 3				< 0.001	$p < 0.001$
			sess. 2 vs. 3				0.006	$p < 0.01$
4	A left	L2/3-D	sess. 1 to null		10^5	24	< 0.001	$p < 0.001$
			sess. 2 to null		permutations	comparisons	< 0.001	$p < 0.001$
			sess. 3 to null				0.084	n.s.
		L2/3-S	sess. 1 to null				< 0.001	$p < 0.001$
			sess. 2 to null				0.069	n.s.
			sess. 3 to null				0.029	$p < 0.05$

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

Fig.	Panel	Comparison		Test	Bonferroni correction	Corrected p-value	Signif.
(4)	(A left)	L5-D	sess. 1 to null	(10^5 permutations)	(24 comparisons)	0.002	$p < 0.01$
			sess. 2 to null			0.010[9]	$p < 0.01$
			sess. 3 to null			0.009	$p < 0.01$
		L5-S	sess. 1 to null			< 0.001	$p < 0.001$
			sess. 2 to null			1.000	n.s.
			sess. 3 to null			0.005	$p < 0.01$
	A right	L2/3-D	sess. 1 to null			0.504	n.s.
			sess. 2 to null			< 0.001	$p < 0.001$
			sess. 3 to null			0.001	$p < 0.01$
		L2/3-S	sess. 1 to null			< 0.001	$p < 0.001$
			sess. 2 to null			0.177	n.s.
			sess. 3 to null			0.667	n.s.
		L5-D	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-S	sess. 1 to null			< 0.001	$p < 0.001$
			sess. 2 to null			0.160	n.s.
			sess. 3 to null			1.000	n.s.
	C	L2/3-D	sess. 1 v 2 to null	10^5 permutations	8 comparisons	0.647	n.s.
			sess. 2 v 3 to null			< 0.001	$p < 0.001$
		L5-D	sess. 1 v 2 to null			0.446	n.s.
			sess. 2 v 3 to null			0.018	$p < 0.05$
		L2/3-S	sess. 1 v 2 to null			< 0.001	$p < 0.001$
			sess. 2 v 3 to null			1.000	n.s.
		L5-S	sess. 1 v 2 to null			0.029	$p < 0.05$
			sess. 2 v 3 to null			1.000	n.s.
S4	A left	L2/3-D	∼6%	binomial null CI	22 comparisons	< 0.001	$p < 0.001$
			∼24%			< 0.001	$p < 0.001$
		L2/3-S	∼5%			0.402	n.s.
			∼6%			0.926	n.s.
			∼46%			< 0.001	$p < 0.001$
		L5-D	∼1%			< 0.001	$p < 0.001$
			∼6%			0.189	n.s.
			∼24%			< 0.001	$p < 0.001$
		L5-S	∼6%			1.000	n.s.
			∼7%			0.660	n.s.
			∼13%			0.499	n.s.
	A right	L2/3-D	∼2%			1.000	n.s.
			∼10%			< 0.001	$p < 0.001$
		L2/3-S	∼3%			1.000	n.s.
			∼9%			< 0.001	$p < 0.001$
			∼51%			< 0.001	$p < 0.001$

[9] Although the corrected p-value is rounded here to 0.010, the actual value is < 0.01.

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

Fig.	Panel	Comparison			Test	Bonferroni correction	Corrected p-value	Signif.
(S4)	(A right)	L5-D	∼1%		(binomial null CI)	(22 comparisons)	0.404	n.s.
			∼6%				0.064	n.s.
			∼10%				< 0.001	$p < 0.001$
		L5-S	∼7%				1.000	n.s.
			∼32%				< 0.001	$p < 0.001$
			∼34%				< 0.001	$p < 0.001$
S4	B	L2/3-D	sess. 1 vs. 2	reg.	10^5 permutations	24 comparisons	1.000	n.s.
				unexp.			< 0.001	$p < 0.001$
			sess. 1 vs. 3	reg.			1.000	n.s.
				unexp.			< 0.001	$p < 0.001$
			sess. 2 vs. 3	reg.			1.000	n.s.
				unexp.			1.000	n.s.
		L2/3-S	sess. 1 vs. 2	reg.			1.000	n.s.
				unexp.			1.000	n.s.
			sess. 1 vs. 3	reg.			0.454	n.s.
				unexp.			1.000	n.s.
			sess. 2 vs. 3	reg.			1.000	n.s.
				unexp.			1.000	n.s.
		L5-D	sess. 1 vs. 2	reg.			1.000	n.s.
				unexp.			< 0.001	$p < 0.001$
			sess. 1 vs. 3	reg.			< 0.001	$p < 0.001$
				unexp.			< 0.001	$p < 0.001$
			sess. 2 vs. 3	reg.			0.045	$p < 0.05$
				unexp.			0.085	n.s.
		L5-S	sess. 1 vs. 2	reg.			1.000	n.s.
				unexp.			0.956	n.s.
			sess. 1 vs. 3	reg.			1.000	n.s.
				unexp.			1.000	n.s.
			sess. 2 vs. 3	reg.			1.000	n.s.
				unexp.			1.000	n.s.
S5	C	L2/3-D	sess. 1 to null		10^5 permutations	24 comparisons	< 0.001	$p < 0.001$
			sess. 2 to null				< 0.001	$p < 0.001$
			sess. 3 to null				< 0.001	$p < 0.001$
			sess. 1 vs. 2				1.000	n.s.
			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				0.003	$p < 0.01$
			sess. 2 to null				0.022	$p < 0.05$
			sess. 3 to null				< 0.001	$p < 0.001$
			sess. 1 vs. 2				1.000	n.s.
			sess. 1 vs. 3				0.245	n.s.
			sess. 2 vs. 3				1.000	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.
(S5)	(C)	L5-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				1.000	n.s.
			sess. 1 vs. 2				< 0.001	$p < 0.001$
			sess. 1 vs. 3				< 0.001	$p < 0.001$
			sess. 2 vs. 3				< 0.001	$p < 0.001$
		L5-S	sess. 1 to null				1.000	n.s.
			sess. 2 to null				1.000	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				1.000	n.s.
S6	A	L2/3-D	sess. 1 vs. 2	exp.	10^5	24	0.127	n.s.
				unexp.	permutations	comparisons	1.000	n.s.
			sess. 1 vs. 3	exp.			1.000	n.s.
				unexp.			< 0.001	$p < 0.001$
			sess. 2 vs. 3	exp.			1.000	n.s.
				unexp.			0.015	$p < 0.05$
		L2/3-S	sess. 1 vs. 2	exp.			1.000	n.s.
				unexp.			1.000	n.s.
			sess. 1 vs. 3	exp.			1.000	n.s.
				unexp.			1.000	n.s.
			sess. 2 vs. 3	exp.			1.000	n.s.
				unexp.			1.000	n.s.
		L5-D	sess. 1 vs. 2	exp.			0.002	$p < 0.01$
				unexp.			0.029	$p < 0.05$
			sess. 1 vs. 3	exp.			0.038	$p < 0.05$
				unexp.			1.000	n.s.
			sess. 2 vs. 3	exp.			1.000	n.s.
				unexp.			1.000	n.s.
		L5-S	sess. 1 vs. 2	exp.			1.000	n.s.
				unexp.			1.000	n.s.
			sess. 1 vs. 3	exp.			1.000	n.s.
				unexp.			1.000	n.s.
			sess. 2 vs. 3	exp.			1.000	n.s.
				unexp.			0.387	n.s.
S6	B	L2/3-D	gabors vs. vis. flow		10^5	5	< 0.001	$p < 0.001$
		L2/3-S	gabors vs. vis. flow		permutations	comparisons	0.004	$p < 0.01$
		L5-D	gabors vs. vis. flow				< 0.001	$p < 0.001$
		L5-S	gabors vs. vis. flow				1.000	n.s.
		all	gabors vs. vis. flow				< 0.001	$p < 0.001$

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

Fig.	Panel	Comparison		Test	Bonferroni correction	Corrected p-value	Signif.
S6	E	L2/3-D	gabors vs. vis. flow	10^5	5	1.000	n.s.
		L2/3-S	gabors vs. vis. flow	permutations	comparisons	< 0.001	$p < 0.001$
		L5-D	gabors vs. vis. flow			< 0.001	$p < 0.001$
		L5-S	gabors vs. vis. flow			0.690	n.s.
		all	gabors vs. vis. flow			0.905	n.s.
S6	F	L2/3-D	sess. 1 v 2 to null	10^5	8	1.000	n.s.
			sess. 2 v 3 to null	permutations	comparisons	0.106	n.s.
		L5-D	sess. 1 v 2 to null			1.000	n.s.
			sess. 2 v 3 to null			1.000	n.s.
		L2/3-S	sess. 1 v 2 to null			1.000	n.s.
			sess. 2 v 3 to null			1.000	n.s.
		L5-S	sess. 1 v 2 to null			1.000	n.s.
			sess. 2 v 3 to null			1.000	n.s.