Report

Current Biology

Predictions and errors are distinctly represented across V1 layers

Highlights

- A 7T fMRI study presents expected and unexpected Gabor orientations
- Expectation differentially modulates decoding across primary visual cortex layers
- This pattern supports predictive processing accounts of the brain and mind

Authors

Emily R. Thomas, Joost Haarsma, Jessica Nicholson, Daniel Yon, Peter Kok, Clare Press

Correspondence

emilyrosethomas@outlook.com (E.R.T.), c.press@ucl.ac.uk (C.P.)

In brief

Thomas et al. use 7T fMRI to find that expected events are represented similarly across deep, middle, and superficial layers of the primary visual cortex, while unexpected events are only robustly represented in superficial layers. These findings support accounts of sensory processing requiring distinct representations of predictions and errors.

Thomas et al., 2024, Current Biology *34*, 1–7
May 20, 2024
https://doi.org/10.1016/j.cub.2024.04.036

CellPress

Current Biology

CellPress
OPEN ACCESS

Report

Predictions and errors are distinctly represented across V1 layers

Emily R. Thomas,[1,2,*] Joost Haarsma,[3] Jessica Nicholson,[2] Daniel Yon,[2] Peter Kok,[3] and Clare Press[2,3,4,5,6,*]
[1]Neuroscience Institute, New York University Medical Center, 435 East 30th Street, New York 10016, USA
[2]Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, UK
[3]Wellcome Centre for Human Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London WC1N 3AR, UK
[4]Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, UK
[5]X (formerly Twitter): @ClarePress
[6]Lead contact
*Correspondence: emilyrosethomas@outlook.com (E.R.T.), c.press@ucl.ac.uk (C.P.)
https://doi.org/10.1016/j.cub.2024.04.036

SUMMARY

Popular accounts of mind and brain propose that the brain continuously forms predictions about future sensory inputs and combines predictions with inputs to determine what we perceive.[1–6] Under "predictive processing" schemes, such integration is supported by the hierarchical organization of the cortex, whereby feedback connections communicate predictions from higher-level deep layers to agranular (superficial and deep) lower-level layers.[7–10] Predictions are compared with input to compute the "prediction error," which is transmitted up the hierarchy from superficial layers of lower cortical regions to the middle layers of higher areas, to update higher-level predictions until errors are reconciled.[11–15] In the primary visual cortex (V1), predictions have thereby been proposed to influence representations in deep layers while error signals may be computed in superficial layers. Despite the framework's popularity, there is little evidence for these functional distinctions because, to our knowledge, unexpected sensory events have not previously been presented in human laminar paradigms to contrast against expected events. To this end, this 7T fMRI study contrasted V1 responses to expected (75% likely) and unexpected (25%) Gabor orientations. Multivariate decoding analyses revealed an interaction between expectation and layer, such that expected events could be decoded with comparable accuracy across layers, while unexpected events could only be decoded in superficial laminae. Although these results are in line with these accounts that have been popular for decades, such distinctions have not previously been demonstrated in humans. We discuss how both prediction and error processes may operate together to shape our unitary perceptual experiences.

RESULTS

Due to previous work, partly from our group,[16,17] demonstrating fast and robust learning of action-cue-outcome relationships in human participants, we used an action-outcome paradigm to establish predictions. Twenty-two participants (17 female, mean age = 26.09 years, and SD = 3.41) were trained with perfect relationships between finger actions and visual Gabor orientations (e.g., index finger abduction = clockwise-oriented Gabor [CW]; little finger abduction = counter-clockwise Gabor [CCW]). They were presented at test (scanning phase, the following day) with degraded contingencies to measure neural responses to "expected" (in line with perfect contingency training phase; 75% of trials in the scanner) and "unexpected" (25%) events (see Figure 1). On half the trials they were asked to give a yes/no response to whether the stimulus was oriented CW and on the other half they were asked whether it was oriented CCW. This design orthogonalized the Gabor orientation presentation from the response. Linear support vector machines (SVMs) were trained to discriminate Gabor orientations from V1 activation during a localizer and were tested on the main task,[18] separately for expected and unexpected events (Figure 2). Under the predictive processing account outlined above, an interaction is hypothesized between expectation and layer in decoding accuracy.

Influence of expectations on behavior

Reaction time (RT) data were collected for responses to expected and unexpected stimuli in the test session and median RTs were calculated for correct trials, separately for each condition and participant. Similarly, the proportion of correct responses was analyzed for expected and unexpected conditions. RT analyses revealed no difference between expected (M = 586.78 ms, SD = 73.42) and unexpected (M = 589.98 ms, SD = 75.60) trials ($t(19) = -0.57$, $p = 0.58$, and $d = -0.13$). Participants were, however, more accurate on expected ($M = 0.97$, $SD = 0.03$) than unexpected (M = 0.95, SD = 0.04) trials ($t(19) = 2.67$, p = 0.015, d = 0.60; see Figure 3A).

Distinct cortical representations of predictions and errors

Using ultra-high-field 7T fMRI (spatial resolution: 0.8 mm isotropic), we examined the brain activity patterns across

Please cite this article in press as: Thomas et al., Predictions and errors are distinctly represented across V1 layers, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.04.036

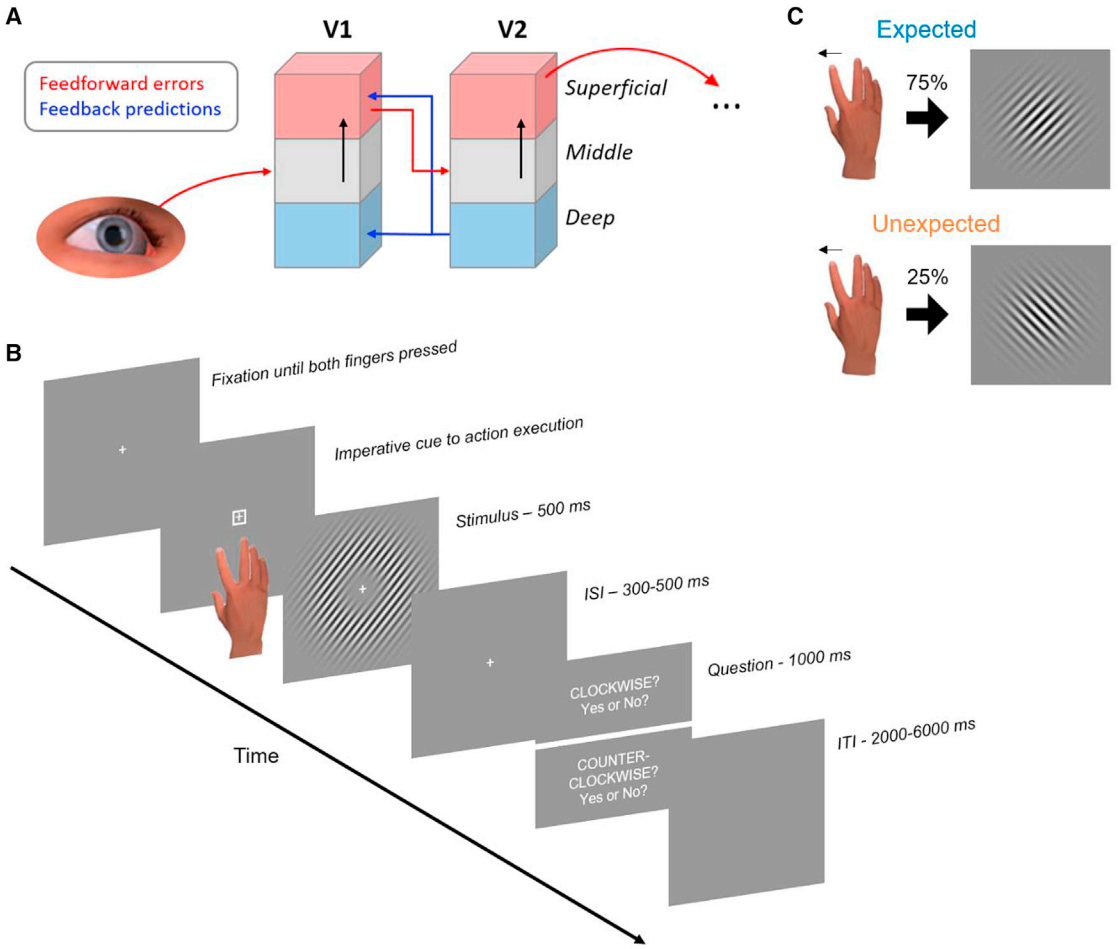

Figure 1. Experiment design

(A) Schematic representation of proposed extrinsic feedforward (red) and feedback (blue) connections across layers in early visual areas.

(B) Experimental paradigm. A centrally presented visual cue instructed participants to abduct either their index or little finger. The imperative cue could be either a triangle or square presented around the fixation cross. Each finger abduction predicted an oriented Gabor and participants were required to respond (yes/no) to whether the stimulus was clockwise (CW) or counter-clockwise (CCW) oriented relative to the vertical. In the training phase, actions perfectly predicted the stimulus orientation (100% contingency). This example demonstrates the relationship for a participant trained in index finger abduction to clockwise-oriented Gabor mappings.

(C) In the scanning session 24 h after the training phase, participants completed the same task, but the action-outcome relationship was degraded to 75% to produce unexpected (25%) as well as expected (75%) events.

cortical layers of the V1 for expected and unexpected Gabor orientations. To determine layer-specific activity patterns, cortical V1 voxels were divided into three equivolume gray matter layer bins—superficial, middle, and deep. The proportion of each voxel's volume across the layers was used to create three cortical layer V1 masks for each participant that were used as layer regions of interest (ROIs) for the following decoding analyses.

To investigate how expectations are represented across the cortical column in the V1, linear SVMs were trained to discriminate Gabor orientations (CW or CCW) from a short localizer task that presented blocks of high-contrast flickering Gabors and were tested on the main task (beta images from a first-level generalized linear model [GLM]). Given that expected events were presented with 75% likelihood, while unexpected events were presented with 25% likelihood, modeling all expected trials would render regressors that contained three times the data of unexpected regressors. We therefore modeled three expected regressors, all with an identical number of trials to unexpected regressors in the GLM, to reduce decoding biases across comparisons. The accuracy scores for each of the three expected decoding conditions were averaged for each participant, providing one accuracy score for expected trials and one for unexpected trials in each of the layer masks. These decoding accuracies were compared using a repeated measures ANOVA.

This analysis revealed a main effect of layer ($F(1.45, 29.04) = 5.98$, $p = 0.012$, $np^2 = 0.23$, Greenhouse-Geisser corrected, $\varepsilon = 0.73$), no main effect of expectation ($F(1, 20) = 0.39$, $p = 0.54$, and $np^2 = 0.019$), and, crucially, an interaction between expectation and layer ($F(2, 40) = 4.45$, $p = 0.018$, and $np^2 = 0.18$; see Figure 3B). Two control analyses were run at the voxel selection stage to balance the number of voxels in each layer mask, considering that there were more voxels in the superficial

Please cite this article in press as: Thomas et al., Predictions and errors are distinctly represented across V1 layers, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.04.036

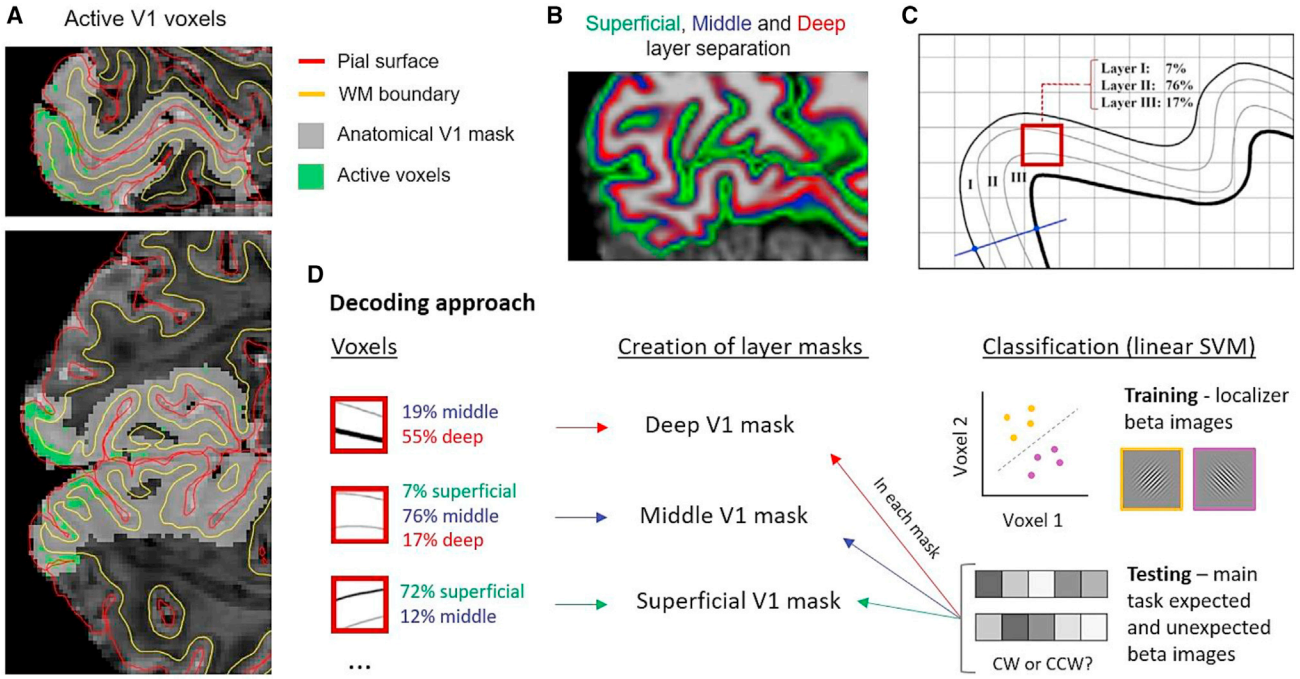

Figure 2. Data analysis

(A) Visualization of the selected anatomical V1 ROI (light gray) on a mean functional image of an example participant. Overlaid red and yellow lines represent co-registered anatomical WM (yellow) and pial surface (red) boundaries to the mean functional image, showing voxels that were significantly active against baseline to the presented stimuli in the functional localizer task (green).

(B) A mean functional image overlaid with distributions of voxels in superficial (green), middle (blue), and deep (red) layers of the cortex.

(C) A schematic representing the level-set approach used to determine the volume distribution of a selected voxel (e.g., red square) over the superficial, middle, and deep cortical layers.[19,20]

(D) A schematic of the decoding approach adopted here. Voxel proportions across the three layer bins in (C) were used to separate voxels according to the majority layer and formed layer masks for V1. Linear classifiers (SVMs) were trained on CW and CCW stimuli from the localizer task and tested on Gabors from the main task. The procedure was repeated separately for expected and unexpected time courses and in each V1 layer mask.

layers (due to known draining vein biases in gradient-echo echo-planar imaging [EPI][21–25]; see STAR Methods). The effects remained in both a voxel-balanced control *t*-map analysis (this analysis selected an equal number of the most active voxels across the three layer bins; see STAR Methods; expectation × layer: $F(2, 40) = 3.64$, $p = 0.035$, $np^2 = 0.15$; layer: $F(1.60, 31.98) = 4.83$, $p = 0.021$, and $np^2 = 0.20$, Huynh-Feldt corrected, $\varepsilon = 0.80$; expectation: $F(1, 20) = 1.77$, $p = 0.20$, and $np^2 = 0.08$), and a random-sample analysis (selecting an equal number of voxels across layers sampled randomly from each layer; expectation × layer: $F(2, 40) = 4.35$, $p = 0.019$, and $np^2 = 0.18$; layer: $F(2, 40) = 4.70$, $p = 0.015$, and $np^2 = 0.19$; expectation: $F(1, 20) = 2.07$, $p = 0.17$, and $np^2 = 0.09$).

This interaction was generated via relatively consistent decoding across layers for expected events ($F(2, 40) = 0.82$, $p = 0.45$, and $np^2 = 0.04$), while decoding performance differed for unexpected events ($F(2, 40) = 6.90$, $p = 0.003$, $np^2 = 0.26$)—increasing from deep to superficial layers (Figure 3B). These effects are complemented by one-sample t tests demonstrating that decoding of expected events was significantly different from chance across all layers (deep: $t(20) = 2.62$, $p = 0.016$; middle: $t(20) = 4.13$, $p = 0.001$; superficial: $t(20) = 2.94$, $p = 0.008$), while unexpected events could only be decoded in superficial layers (deep: $t(20) = -0.73$, $p = 0.47$; middle: $t(20) = 1.03$, $p = 0.31$; superficial: $t(20) = 3.97$, $p = 0.001$). Additional post hoc tests revealed no significant differences between expected and unexpected decoding within each layer (deep: $t(20) = 2.01$, $p = 0.058$, d = 0.44; middle: $t(20) = 0.94$, $p = 0.36$, d = 0.21; superficial: $t(20) = -1.41$, $p = 0.18$, and d = 0.31), though the numerical differences reveal superior representation of expected events in deep (expected: $M = 3.27$, $SD = 5.73$; unexpected: $M = -1.76$, $SD = 11.21$) and middle layers (expected: $M = 5.36$, $SD = 5.95$; unexpected: $M = 2.98$, $SD = 13.20$), flipping to superior representation of the unexpected in superficial layers (expected: $M = 4.76$, $SD = 7.43$; unexpected: $M = 8.63$, $SD = 9.98$).

Taken together, these tests demonstrate that representation of unexpected events increases toward the superficial layers of the V1, only becoming significantly decodable in these layers, while expected events are represented similarly across the cortical column.

DISCUSSION

This study examined how unexpected visual events are represented across cortical layers, in comparison with expected events, in a high-resolution fMRI study. It found, in line with previous work,[19] that expected events were represented (decoded) equivalently across deep, middle, and superficial bins but, more

Please cite this article in press as: Thomas et al., Predictions and errors are distinctly represented across V1 layers, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.04.036

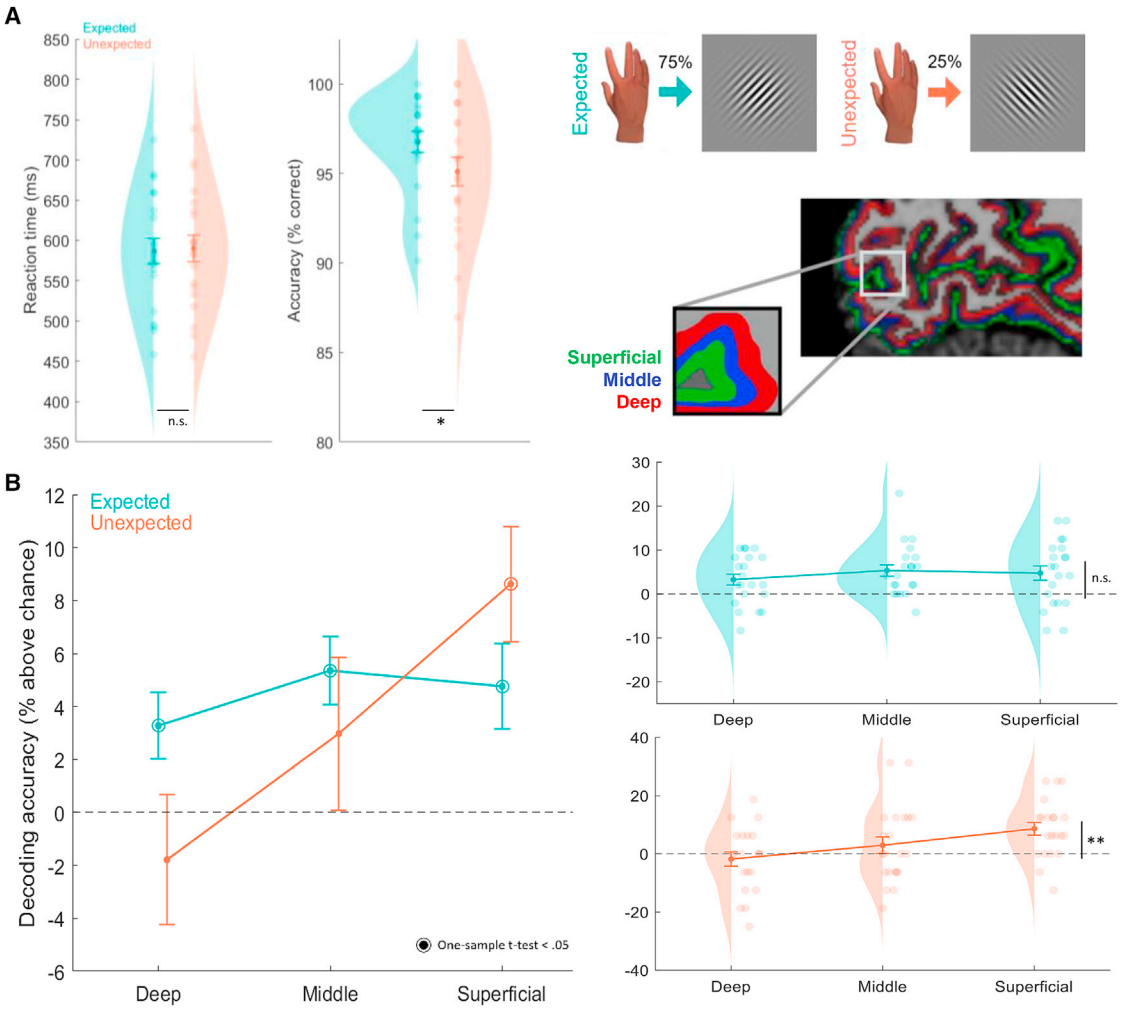

Figure 3. Results

(A) Mean RTs and accuracy (± SEM) for expected and unexpected events alongside probability density estimates and individual participant data points. There was no difference between conditions in RT, but participants were more accurate in expected than unexpected judgments (*$p < 0.05$).

(B) The results of the decoding analysis across cortical layer bins, where mean (± SEM) decoding accuracy percentage above chance (50%) is plotted for expected and unexpected trials. On the left panel, circles around mean data points indicate that the decoding accuracy was significantly above chance ($p < 0.05$), which was the case across all layers for expected events but only in superficial layers for unexpected events. On the right panel, decoding accuracies are plotted alongside probability density estimates and individual participant data points. The linear trend across layers was significant for unexpected (**$p < 0.001$) but not expected events.

novelly, that unexpected events were represented with varying fidelity—such that they were poorly represented in deep and middle layers and could only be decoded above chance in superficial layers.

These findings are in line with predictive processing accounts in which predictions and errors are represented distinctly across cortical laminae, such that predictions conveyed via feedback projections inject input into hypothesis units in deep layers, while feedforward connections transmit the error from the superficial laminae.[12] This finding is also in line with data from mice and monkeys indicating superficial layer discrepancy signals with respect to other types of feedback[26–28] (see also Gillon et al.[29] and Fiser et al.[30]). Although this account of cortical processing has been popular for a couple of decades, such distinctions have not previously been demonstrated in human cortical processing.

It is worth noting that alternatives have been proposed to this account to explain precisely how prediction and input signals are combined in the brain. Particularly, a number of fMRI studies have observed improved sensory decoding of events expected on the basis of preceding cues relative to unexpected events,[16,17,31,32] which may suggest that channels tuned to expected inputs are more responsive than channels tuned to the unexpected. Such a "global sharpening" account[33] proposes that the precision weights are adjusted to increase the gain of expected channels, allowing them to respond more sensitively to expected input and subsequently improving representation of the expected across the cortical column. This account would predict facilitated processing of the expected[1–5] across layers, if we relatively inhibit processing via other channels across layers, and therefore is inconsistent with the patterns observed here.

Please cite this article in press as: Thomas et al., Predictions and errors are distinctly represented across V1 layers, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.04.036

Predictive processing accounts were initially developed to explain inference about the present—removing redundancy in the system—while the difference between our expected and unexpected conditions pertains to cues that allow one to predict statistical likelihoods about *future* events. The fact that we see differences between conditions like these demonstrates that future-based (e.g., cue-based) predictions inject hypotheses into units in deep layers[19] in the same way as present-state predictions. These findings are in line with evidence from human ultra-high resolution 7T MRI studies demonstrating activity in deep cortical layers of the V1 for visual events that are expected but never presented[19,34] (see also Muckli et al.[35]).

Forging functional conclusions about the operation of mechanisms across cortical layers has become possible with high-resolution MRI but is, of course, also plagued by interpretational issues due to venous draining of blood toward the pial surface.[21–25] Specifically, gradient-echo blood-oxygen-level-dependent (BOLD) signal is known to exhibit strong contributions from large veins situated perpendicular to the cortical surface as venous blood is drained from lower to upper cortical layers.[15,22,24] Such venous issues render it likely, for instance, that neural effects at deeper cortical layers contribute to responses in superficial layers.[36] Other labs favor cerebral-blood-volume-based vascular-space-occupancy (VASO)[37] fMRI due to superior laminar separation, although this is accompanied by reduced content-based sensitivity relative to gradient-echo methods.[38] Here, we use gradient-echo BOLD but mitigate such contributions by comparing responses between stimuli that are identical—other than their expectedness due to a preceding cue—because venous draining influences should be equivalent for both expected and unexpected events. Another methodological debate in the field surrounds separation of the signal into three cortical layers,[19,20,36,39–41] such as here, versus more layers. Given the voxel size of 0.8 mm and a cortical thickness of 2.5–3 mm, three layers could be conceived to be the most realistic resolution to be achieved with this voxel size.[42] Nevertheless, importantly for our conclusions, we know of no literature that would suggest such an expectation × layer interaction effect, as observed here, would be generated by our methodological choices and not reflective of true mechanistic differences, but future work would, of course, be wise to investigate replicability with different approaches.

Such distinct representation of prediction and error may be an adaptive solution allowing predictions to shape perception to serve a number of functions. Some of us have recently discussed how predictions often need to exhibit quite distinct behavioral shaping of perception to serve the organism.[4,43] To overcome noise in sensory processing and generate broadly accurate experiences rapidly, we may bias perception toward what we expect.[44,45] However, larger error signals (that cannot have resulted from noise) may require high perceptual resources to enable accurate perception and resultant model updating. If we represent the error signal separate from the prediction, even in early sensory processing, this may be one way to enable these large error signals to communicate deviation rapidly to systems mediating model updating—such as the locus coeruleus.[46] Future work must establish how these error signals relate to perception and model updating to truly test these accounts and examine whether error signals in superficial layers are calculated in the first feedforward sweep[47] or subsequent stimulus-processing iterations.

It has been suggested in various theoretical accounts that symptoms of psychosis, like hallucinations and delusions, can be explained in terms of aberrant signaling of prediction error as well as overweighting of expectations.[48–50] As demonstrated in this study, laminar fMRI is capable of distinguishing the representation of these signals across different cortical layers. Therefore, laminar fMRI would be well suited to test the theoretical predictions from predictive coding models of psychosis, as well as comparing these mechanisms in other clinical and neurological populations characterized by aberrant perceptual inference, like Parkinson's disease.[51]

In conclusion, this study provides evidence that expected and unexpected visual events are distinctly represented across the cortical column in the V1 via a novel 7T fMRI design that presented unexpected visual events alongside expected counterparts. Expected events were represented similarly across layers but unexpected events were only represented well in superficial layers. These findings contribute to our understanding of how predictions can interact with sensory inputs to shape what we perceive and how we interact with the world.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participants
- METHOD DETAILS
  - Stimuli
  - Procedure
  - Image acquisition
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Behavioural analyses
  - fMRI data preprocessing
  - Cortical layer definition
  - Layer-specific ROI definition
  - Decoding analysis

ACKNOWLEDGMENTS

This work was supported by a Leverhulme Trust project grant (RPG-2016-105) and European Research Council (ERC) consolidator grant (101001592) under the European Union's Horizon 2020 research and innovation programme, both awarded to C.P. P.K. was supported by a Wellcome/Royal Society Sir Henry Dale Fellowship (218535/Z/19/Z) and an ERC starting grant (948548). E.R.T. and DY were supported by the Leverhulme Trust grant awarded to C.P. and J.H. by the ERC grant awarded to P.K. The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome Trust (203147/Z/16/Z). We are grateful to Martina Callaghan for useful discussions.

AUTHOR CONTRIBUTIONS

Conceptualization, E.R.T., D.Y., P.K., and C.P.; formal analysis, investigation, and project administration, E.R.T., J.H., and J.N.; writing – original draft, E.R.T.; writing – review and editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 22, 2024
Revised: April 9, 2024
Accepted: April 13, 2024
Published: May 1, 2024

REFERENCES

1. Bar, M. (2004). Visual objects in context. Nat. Rev. Neurosci. *5*, 617–629. https://doi.org/10.1038/nrn1476.
2. Yuille, A., and Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? Trends Cogn. Sci. *10*, 301–308. https://doi.org/10.1016/j.tics.2006.05.002.
3. de Lange, F.P., Heilbron, M., and Kok, P. (2018). How Do Expectations Shape Perception? Trends Cogn. Sci. *22*, 764–779. https://doi.org/10.1016/j.tics.2018.06.002.
4. Press, C., Kok, P., and Yon, D. (2020). The Perceptual Prediction Paradox. Trends Cogn. Sci. *24*, 13–24. https://doi.org/10.1016/J.TICS.2019.11.003.
5. Den Ouden, H.E., Kok, P., and De Lange, F.P. (2012). How Prediction Errors Shape Perception, Attention, and Motivation. Front. Psychol. *3*, 548. https://doi.org/10.3389/fpsyg.2012.00548.
6. Kaiser, D., Quek, G.L., Cichy, R.M., and Peelen, M.V. (2019). Object Vision in a Structured World. Trends Cogn. Sci. *23*, 672–685. https://doi.org/10.1016/j.tics.2019.04.013.
7. Felleman, D.J., and Van Essen, D.C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. Cereb. Cortex *1*, 1–47. https://doi.org/10.1093/cercor/1.1.1-a.
8. Rockland, K.S., and Virga, A. (1989). Terminal arbors of individual "Feedback" axons projecting from area V2 to V1 in the macaque monkey: A study using immunohistochemistry of anterogradely transported Phaseolus vulgaris-leucoagglutinin. J. Comp. Neurol. *285*, 54–72. https://doi.org/10.1002/cne.902850106.
9. van Kerkoerle, T., Self, M.W., and Roelfsema, P.R. (2017). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. Nat. Commun. *8*, 13804. https://doi.org/10.1038/ncomms13804.
10. Yu, Y., Huber, L., Yang, J., Jangraw, D.C., Handwerker, D.A., Molfese, P.J., Chen, G., Ejima, Y., Wu, J., and Bandettini, P.A. (2019). Layer-specific activation of sensory input and predictive feedback in the human primary somatosensory cortex. Sci. Adv. *5*, eaav9053. https://doi.org/10.1126/sciadv.aav9053.
11. Friston, K. (2005). A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. *360*, 815–836. https://doi.org/10.1098/rstb.2005.1622.
12. Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical Microcircuits for Predictive Coding. Neuron *76*, 695–711. https://doi.org/10.1016/j.neuron.2012.10.038.
13. Kanai, R., Komura, Y., Shipp, S., and Friston, K. (2015). Cerebral hierarchies: Predictive processing, precision and the pulvinar. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*. https://doi.org/10.1098/rstb.2014.0169.
14. Rao, R.P.N., and Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. *2*, 79–87. https://doi.org/10.1038/4580.
15. Stephan, K.E., Petzschner, F.H., Kasper, L., Bayer, J., Wellstein, K.V., Stefanics, G., Pruessmann, K.P., and Heinzle, J. (2019). Laminar fMRI and computational theories of brain function. NeuroImage *197*, 699–706. https://doi.org/10.1016/j.neuroimage.2017.11.001.
16. Yon, D., Thomas, E.R., Gilbert, S.J., de Lange, F.P., Kok, P., and Press, C. (2023). Stubborn Predictions in Primary Visual Cortex. J. Cogn. Neurosci. *35*, 1133–1143. https://doi.org/10.1162/jocn_a_01997.
17. Yon, D., Gilbert, S.J., De Lange, F.P., and Press, C. (2018). Action sharpens sensory representations of expected outcomes. Nat. Commun. *9*, 4288. https://doi.org/10.1038/s41467-018-06752-7.
18. Peelen, M., and Downing, P. (2023). Testing cognitive theories using multivariate pattern analysis of neuroimaging data. Nat Hum Behav. *7*, 1430–1441. https://doi.org/10.1038/s41562-023-01680-z.
19. Aitken, F., Menelaou, G., Warrington, O., Koolschijn, R.S., Corbin, N., Callaghan, M.F., and Kok, P. (2020). Prior expectations evoke stimulus-specific activity in the deep layers of the primary visual cortex. PLoS Biol. *18*, e3001023. https://doi.org/10.1371/journal.pbio.3001023.
20. van Mourik, T., van der Eerden, J.P.J.M., Bazin, P.L., and Norris, D.G. (2019). Laminar signal extraction over extended cortical areas by means of a spatial GLM. PLOS ONE *14*, e0212493. https://doi.org/10.1371/journal.pone.0212493.
21. Lawrence, S.J.D., Formisano, E., Muckli, L., and De Lange, F.P. (2019). Laminar fMRI: Applications for cognitive neuroscience. NeuroImage *197*, 785–791. https://doi.org/10.1016/j.neuroimage.2017.07.004.
22. Duvernoy, H.M., Delon, S., and Vannson, J.L. (1981). Cortical blood vessels of the human brain. Brain Res. Bull. *7*, 519–579. https://doi.org/10.1016/0361-9230(81)90007-1.
23. Koopmans, P.J., Barth, M., and Norris, D.G. (2010). Layer-specific BOLD activation in human V1. Hum. Brain Mapp. *31*, 1297–1304. https://doi.org/10.1002/hbm.20936.
24. Markuerkiaga, I., Barth, M., and Norris, D.G. (2016). A cortical vascular model for examining the specificity of the laminar BOLD signal. NeuroImage *132*, 491–498. https://doi.org/10.1016/j.neuroimage.2016.02.073.
25. Heinzle, J., Koopmans, P.J., den Ouden, H.E.M., Raman, S., and Stephan, K.E. (2016). A hemodynamic model for layered BOLD signals. NeuroImage *125*, 556–570. https://doi.org/10.1016/j.neuroimage.2015.10.025.
26. Audette, N.J., Zhou, W., La Chioma, A., and Schneider, D.M. (2022). Precise movement-based predictions in the mouse auditory cortex. Curr. Biol. *32*, 4925–4940.e6. https://doi.org/10.1016/j.cub.2022.09.064.
27. 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 *108*, 1194–1206.e5. https://doi.org/10.1016/j.neuron.2020.09.024.
28. Bastos, A.M., Lundqvist, M., Waite, A.S., Kopell, N., and Miller, E.K. (2020). Layer and rhythm specificity for predictive routing. Proc. Natl. Acad. Sci. USA *117*, 31459–31469. https://doi.org/10.1073/pnas.2014868117.
29. Gillon, C.J., Pina, J.E., Lecoq, J.A., Ahmed, R., Billeh, Y.N., Caldejon, S., Groblewski, P., Henley, T.M., Kato, I., Lee, E., et al. (2023). Learning from unexpected events in the neocortical microcircuit. Preprint at bioRxiv. https://doi.org/10.1101/2021.01.15.426915.
30. Fiser, A., Mahringer, D., Oyibo, H.K., Petersen, A.V., Leinweber, M., and Keller, G.B. (2016). Experience-dependent spatial expectations in mouse visual cortex. Nat. Neurosci. *19*, 1658–1664. https://doi.org/10.1038/nn.4385.
31. Kok, P., Jehee, J.F.M., and de Lange, F.P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. Neuron *75*, 265–270. https://doi.org/10.1016/j.neuron.2012.04.034.
32. Brandman, T., and Peelen, M.V. (2023). Objects sharpen visual scene representations: evidence from MEG decoding. Cereb. Cortex *33*, 9524–9531. https://doi.org/10.1093/cercor/bhad222.
33. Friston, K. (2018). Does predictive coding have a future? Nat. Neurosci. *21*, 1019–1021. https://doi.org/10.1038/s41593-018-0200-7.
34. Haarsma, J., Deveci, N., Corbin, N., Callaghan, M.F., and Kok, P. (2023). Perceptual expectations and false percepts generate stimulus-specific activity in distinct layers of the early visual cortex. J Neurosci. *43*, 7946–7957. https://doi.org/10.1523/JNEUROSCI.0998-23.2023.
35. Muckli, L., De Martino, F., Vizioli, L., Petro, L.S., Smith, F.W., Ugurbil, K., Goebel, R., and Yacoub, E. (2015). Contextual Feedback to Superficial Layers of V1. Curr. Biol. *25*, 2690–2695. https://doi.org/10.1016/j.cub.2015.08.057.

36. Kok, P., Bains, L.J., van Mourik, T., Norris, D.G., and de Lange, F.P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. Curr. Biol. *26*, 371–376. https://doi.org/10.1016/j.cub.2015.12.038.

37. Lu, H., Golay, X., Pekar, J.J., and Van Zijl, P.C.M. (2003). Functional magnetic resonance imaging based on changes in vascular space occupancy. Magn. Reson. Med. *50*, 263–274. https://doi.org/10.1002/mrm.10519.

38. Huber, L., Goense, J., Kennerley, A.J., Trampel, R., Guidi, M., Reimer, E., Ivanov, D., Neef, N., Gauthier, C.J., Turner, R., and Möller, H.E. (2015). Cortical lamina-dependent blood volume changes in human brain at 7 T. NeuroImage *107*, 23–33. https://doi.org/10.1016/j.neuroimage.2014.11.046.

39. Lawrence, S.J.D., van Mourik, T., Kok, P., Koopmans, P.J., Norris, D.G., and de Lange, F.P. (2018). Laminar Organization of Working Memory Signals in Human Visual Cortex. Curr. Biol. *28*, 3435–3440.e4. https://doi.org/10.1016/j.cub.2018.08.043.

40. Lawrence, S.J., Norris, D.G., and de Lange, F.P. (2019). Dissociable laminar profiles of concurrent bottom-up and top-down modulation in the human visual cortex. eLife *8*, e44422. https://doi.org/10.7554/eLife.44422.

41. Haarsma, J., Deveci, N., Corbin, N., Callaghan, M.F., and Kok, P. (2022). Perceptual expectations and false percepts generate stimulus-specific activity in distinct layers of the early visual cortex. J Neurosci. *43*, 7946–7957. https://doi.org/10.1523/JNEUROSCI.0998-23.2023.

42. de Sousa, A.A., Sherwood, C.C., Schleicher, A., Amunts, K., MacLeod, C.E., Hof, P.R., and Zilles, K. (2010). Comparative Cytoarchitectural Analyses of Striate and Extrastriate Areas in Hominoids. Cereb. Cortex *20*, 966–981. https://doi.org/10.1093/cercor/bhp158.

43. Press, C., Kok, P., and Yon, D. (2020). Learning to Perceive and Perceiving to Learn. Trends Cogn. Sci. *24*, 260–261. https://doi.org/10.1016/j.tics.2020.01.002.

44. Thomas, E.R., Yon, D., De Lange, F.P., and Press, C. (2022). Action Enhances Predicted Touch. Psychol. Sci. *33*, 48–59. https://doi.org/10.1177/09567976211017505.

45. Press, C., and Yon, D. (2019). Perceptual Prediction: Rapidly Making Sense of a Noisy World. Curr. Biol. *29*, R751–R753. https://doi.org/10.1016/j.cub.2019.06.054.

46. Yu, A.J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. Neuron *46*, 681–692. https://doi.org/10.1016/j.neuron.2005.04.026.

47. Alilović, J., Timmermans, B., Reteig, L.C., van Gaal, S., and Slagter, H.A. (2019). No Evidence that Predictions and Attention Modulate the First Feedforward Sweep of Cortical Information Processing. Cereb. Cortex *29*, 2261–2278. https://doi.org/10.1093/cercor/bhz038.

48. Randeniya, R., Oestreich, L.K.L., and Garrido, M.I. (2018). Sensory prediction errors in the continuum of psychosis. Schizophr. Res. *191*, 109–122. https://doi.org/10.1016/j.schres.2017.04.019.

49. Sterzer, P., Adams, R.A., Fletcher, P., Frith, C., Lawrie, S.M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., and Corlett, P.R. (2018). The Predictive Coding Account of Psychosis. Biol. Psychiatry *84*, 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015.

50. Kafadar, E., Fisher, V.L., Quagan, B., Hammer, A., Jaeger, H., Mourgues, C., Thomas, R., Chen, L., Imtiaz, A., Sibarium, E., et al. (2022). Conditioned Hallucinations and Prior Overweighting Are State-Sensitive Markers of Hallucination Susceptibility. Biol. Psychiatry *92*, 772–780. https://doi.org/10.1016/j.biopsych.2022.05.007.

51. Haarsma, J., Kok, P., and Browning, M. (2022). The promise of layer-specific neuroimaging for testing predictive coding theories of psychosis. Schizophr. Res. *245*, 68–76. https://doi.org/10.1016/j.schres.2020.10.009.

52. Greve, D.N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. NeuroImage *48*, 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060.

53. van Mourik, T., Koopmans, P.J., and Norris, D.G. (2019). Improved cortical boundary registration for locally distorted fMRI scans. PLOS ONE *14*, e0223440. https://doi.org/10.1371/journal.pone.0223440.

54. Hebart, M.N., Görgen, K., and Haynes, J.D. (2014). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. Front. Neuroinform. *8*, 88. https://doi.org/10.3389/fninf.2014.00088.

55. He, H., and Garcia, E.A. (2009). Learning from Imbalanced Data. IEEE Trans. Knowl. Data Eng. *21*, 1263–1284. https://doi.org/10.1109/TKDE.2008.239.

Please cite this article in press as: Thomas et al., Predictions and errors are distinctly represented across V1 layers, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.04.036

STAR★METHODS

KEY RESOURCES TABLE

Deposited information	Source	Identifier
Analysis code and data	This paper	DOI: http://osf.io/s2z8c

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Clare Press (c.press@ucl.ac.uk).

Materials availability

This study did not generate any new materials.

Data and code availability

- The GLM-generated beta data will be deposited at OSF and will be publicly available as of the date of publication. DOIs are listed in the key resources table.
- All original code will be deposited at OSF and will be publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to re-analyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants

Twenty-two participants (17 female, mean age = 26.09 years, SD = 3.41) were recruited from UCL and Birkbeck, University of London, and paid a small honorarium for participation. All participants reported normal or corrected to normal vision and had no history of psychiatric or neurological illness. We did not analyse the effects according to demographic differences, because we only collected gender and age information, and there was no plan to conduct such analyses at any point. In principle this limits the generalisability of our findings, but there is no evidence from previous work to suggest such low level and fundamental visual functions would differ according to these characteristics. One participant's data were excluded due to a technical error during acquisition, which meant that event onsets in one run could not be modelled. This resulted in a final sample of 21 participants. The experiment was approved by the UCL ethics committee.

METHOD DETAILS

Stimuli

Sinusoidal grating (Gabor) stimuli were created using MATLAB and presented against a grey background using Cogent Graphics. During pre-scanner training, stimuli were presented on a 14" LCD screen (resolution: 1280x1024; refresh rate: 60 Hz) at a viewing distance of 45 cm, and during scanning on an LCD monitor (resolution: 1280x1024; refresh rate: 60 Hz) through a mirror at a viewing distance of 91 cm. In both sessions, stimuli were viewed at 15 degrees of visual angle. A Gaussian filter enveloped the grating stimuli to create Gabor patches of 80% Michelson contrast, at 1.5 cycles per degree, and with random spatial phase. The Gabor stimuli were presented in an annulus around a fixation cross in the middle of the screen (see Figure 1B). Two stimulus orientations were generated to appear in CW (45°) and CCW (135°) orientations (relative to the hypothetical vertical mid-point e.g., 90°).

Procedure

Main task

Participants completed two sessions. First, they completed a training session in which finger abductions perfectly predicted visually presented Gabor orientations. The following day, they completed the same task in the MRI scanner but the action-outcome relationship was degraded to 75% validity to allow for presentation of unexpected (25%) as well as expected events.

Participants completed the training session on Gorilla (www.gorilla.sc) for online experiments, taking part on either a laptop or desktop computer no more than 24 hrs before the scanning session. Instruction at the beginning of the experiment requested participants to set screen brightness to the maximum level to reduce variability in viewing conditions. Each trial started with a white

Please cite this article in press as: Thomas et al., Predictions and errors are distinctly represented across V1 layers, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.04.036

fixation cross. Participants were instructed to depress the 'c' and 'm' computer keys with their right index and little fingers, respectively, until an imperative cue (e.g., square or triangle overlaid around the fixation cross) indicated which finger to abduct. A right-hand index finger abduction would involve the finger moving left of hand midline, while a right hand little finger abduction moves right of the midline. After the appropriate action was executed, the imperative cue was replaced with an oriented Gabor for 500 ms, resulting in apparent synchrony of stimulus onset with action execution. A variable 300 – 500 ms delay followed stimulus offset and preceded a response screen which asked about Gabor orientation. On half the trials they were asked to give a yes/no response to whether the stimulus was oriented CW and on the other half they were asked whether it was oriented CCW. This design orthogonalised the Gabor from the response. Participants were required to respond to the question screen within 1500 ms and the next trial started after a variable ITI of 2000-3000 ms. Responses were made using the left thumb on the 'a' and 'z' keys for 'yes' or 'no', respectively. The response question alternated every block. Participants completed the training task in ten runs of 36 trials each.

The following day, participants completed the test session at the Wellcome Centre for Human Neuroimaging, UCL. The test session task was largely similar to the training session except that participants' abductions now predicted the stimulus orientation with 75% validity and they performed actions using MR-compatible button boxes instead of the keyboard. The right-hand button box was positioned orthogonally to the screen in the scanner, in line with the body midline. A short refresher of the training session was presented immediately before the scanning session, using the MR compatible button boxes outside of the scanner. Responses were now required within 1000 ms of the question screen (to reduce scanning time), and the response question was randomly selected on each trial. The next trial started after a variable ITI of 2000-6000 ms. Participants completed the test session in four scanning runs that contained 96 trials each, and a 30 s break was presented mid-way through each run.

There were 384 main experimental trials in the scanning session, 360 online training trials and 192 refresher training trials. This number of trials was determined based on a preceding 3T fMRI study using a comparable task design.[16] Participants completed 32 practice trials before proceeding to the main trials in the initial training session. Imperative cue order and trial order were randomised within blocks and the specific action-Gabor (predictive) relationship was counterbalanced across participants. The imperative cue-action mapping was also counterbalanced and reversed halfway through each session (e.g., at the beginning of the sixth block in training, and beginning of the third block in scanning) to deconfound potential influences of cue-outcome learning and remove any correlation between the imperative action cues and actual or expected Gabor orientations across the experiment.

Localiser task

At the end of the main experiment, participants completed a functional localiser task in an additional scanning run. This task presented flickering Gabor stimuli at approximately 1.8 Hz along with a fixation cross. These Gabors were identical to those presented in the main experiment except that they were presented at 100% contrast, and in blocks of 14 s. Each block containing flickering Gabors was followed by a blank screen containing only the fixation cross for the same duration. In each stimulus block, Gabor orientation was either CW or CCW and the presentation order was pseudorandomised. The task required participants to respond by pressing any button when the central fixation cross changed colour from white to grey, ensuring that their fixation remained central. In total, 32 blocks of flickering Gabors were presented, 16 of each orientation.

Image acquisition

Images were acquired using a 7T Magnetom MRI scanner (Siemens Healthcare GmbH, Erlangen, Germany) using a 32-channel head coil at the Wellcome Centre for Human Neuroimaging, UCL. Functional images were acquired using T2*-weighted 3D gradient-echo EPI sequence (3,552 ms volume acquisition time, TR = 74 ms, TE = 26.95 ms, 48 slices, 15° flip angle, voxel size: 0.8 x 0.8 x 0.8 mm, field of view: 192 x 192 x 39 mm). Structural images were acquired using a Magnetization Prepared Two Rapid Acquisition Gradient Echo (MP2RAGE) sequence (TR = 5,000 ms, TE = 2.60 ms, TI = 900 ms, 240 slices, voxel size 0.7 × 0.7 × 0.7 mm, 5° flip angle, field of view 208 × 208 × 156 mm).

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioural analyses

RT data were collected for responses to expected and unexpected stimuli in the test session and median RTs were calculated for correct trials separately for each condition, for each participant. Similarly, the proportion of correct responses was analysed for expected and unexpected conditions for each participant. One participant was removed from the behavioural analysis due to missing almost half of the responses (44% of trials) and performing similarly to chance on the remainder (62% accuracy; note that this participant was maintained for the imaging analysis, but the significance patterns were identical if they were removed).

fMRI data preprocessing

Preprocessing of the images was conducted in SPM12 and Freesurfer (http://surfer.nmr.mgh.harvard.edu/). Functional images were cropped to select only the occipital lobe, to account for distortions in the frontal lobes. These cropped functional images were spatially realigned to the mean image within runs, but also across runs. The temporal signal-to-noise ratio (tSNR, defined as mean signal/SD over time) was calculated before and after spatial realignment and was found to be significantly higher after (M =14.31, SD = 1.23) than before (M = 10.21, SD = 1.05) realignment ($t(20) = -22.10$, $p < .001$).

The realigned functional images were co-registered to the cortical surfaces estimated in participants' MP2RAGE scans in several steps. First, boundaries between grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) were detected using Freesurfer

Please cite this article in press as: Thomas et al., Predictions and errors are distinctly represented across V1 layers, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.04.036

on re-constructed structural scans (skull removed) and were manually corrected to remove any dura that was inaccurately classified as part of the GM surface. A rigid body boundary-based registration (BBR)[52] was used to register GM boundaries to the mean functional image, and a further recursive boundary-based registration (RBR)[53] applied the BBR recursively to portions of cortical mesh in 6 iterations.

Cortical layer definition

The level set method was used to divide GM into three equivolume layers for cortical layer definition (for details see[20]). This method was used to separate five cortical bins (3 GM, WM and CSF) and determine three GM layers (deep, middle, and superficial) by calculating two intermediate surfaces between the WM and pial boundaries. In human V1, these three layer bins have been suggested to correspond to histological layers 1 to 3, layer 4, and layers 5 and 6, respectively.[19]

Layer-specific ROI definition

Freesurfer was used to define V1 based on anatomical landmarks in the MP2RAGE scans. ROIs were restricted to voxels from the preprocessed functional localiser data that were most active during blocked presentation of the stimuli. This was achieved by modelling regressors for blocks of CW and CCW stimuli against baseline in a temporal GLM to identify voxels that expressed a significant response to these stimuli ($t > 2.3$, $p < 0.05$; $M = 5420.57$, $SD = 2228.12$ number of voxels). A V1 mask of active voxels was created this way for each participant.

Next, these active voxel masks were used to design a matrix of distributed voxels across each layer bin using the level-set definition described earlier.[20] These participant-specific design matrices specified the proportion of each active voxel across the 5 layer bins specified above (3GM, WM, CSF), where each voxel was binned into one of the three GM layer bins according to its majority proportion (Figure 2D; see online materials for a complementary univariate analysis approach). For example, a voxel that was spatially located 7% in superficial, 76% in middle and 17% in deep layers would be labelled as a middle layer voxel and selected to contribute to the voxels in the middle layer mask. An arbitrary threshold was set such that the majority proportion for a voxel to be included in a layer mask was >0.4 (40%). This meant that any voxels with roughly equal proportion in each layer bin would not be selected. Importantly, the results did not change when this threshold was removed, since the majority of voxels' 'winning' proportion was greater than 0.4. Using this approach, three layer masks were created from the active V1 voxels for each participant.

This method of defining layer specific ROIs yields V1 layer masks that differ in the number of voxels that contribute to each layer in each participant. Notably, there is a consistently greater number of voxels in superficial (M = 1613.95, SD = 678.17) than middle (M = 1172.48, SD = 530.94) and deep (M = 907.62, SD = 442.68) layer masks (one-way ANOVA: $F(2,40) = 120.50$, $p < .001$, $np^2 = .86$). We therefore performed another analysis to control for these differences, considering that greater information contributing to the decoding signals in superficial layers relative to the other layers may confound our interpretations. Here, the steps are identical to above, except that an additional step was performed to equalise the number of voxels present in each layer ROI mask. Specifically, we defined the number of voxels to select in each mask as the maximum number common to all layers. For example, if the deep mask had the fewest and contained 831 voxels, 831 voxels would be selected across all layers. Next, we loaded in an orientation preference t-map from the GLM specified above, that contrasted CW and CCW regressors against each other, to select the (e.g., 831) most orientation-tuned voxels from each layer. These voxels were those that contributed to each layer mask, such that each layer mask contained an equivalent number of voxels in each layer. Another control version selected the (e.g. 831) voxels randomly from all the active voxels in each layer. Importantly, the results did not change across these selection methods, suggesting that differences in voxel numbers across layers should not alter interpretation (see Results).

Decoding analysis

Multivariate decoding analyses were implemented using the TDT toolbox[54] in MATLAB. We used a cross-classification approach whereby a linear SVM was trained to discriminate Gabor orientations (CW or CCW) presented during the localizer task. This independent dataset ensured that the trained classifier was not biased with any information about the predictability of stimuli. For this step, we reran the GLM that we used for ROI definition above, but instead specified the onsets for each block in the localizer task as separate regressors. Movement parameters were also modelled as nuisance regressors. This GLM resulted in 16 beta images for each orientation that were fed into the SVM for training.

Next, we specified the test data in our cross-classification decoding approach from our main experimental task data. Specifically, we reran and modified the GLM previously run on the main task data that included separate regressors for each condition type (expected, unexpected) and stimulus type (CW, CCW) in each experiment run, in two ways. First, considering that we only had 4 scanning runs, yet it is well established that decoding data is more reliable with increased number of samples, we modelled each condition according to the first and second halves of each run (since there was a 30s break in between continuous scanning; note also that all trial types were balanced within each run half). This resulted in 8 condition regressors for each scanning run (2x ExpCW1, ExpCCW1, UnexpCW1, UnexpCCW1). Second, we ensured that each modelled regressor would have equal weight in terms of the number of trials contributing to each image, considering known biases in decoding performance with unequal numbers of trials.[55] We therefore modelled expected conditions with the same number of trials as those that contribute to unexpected regressors, by randomly sampling from expected trials to form three different expected regressors (see Results). Again, movement parameters were modelled as nuisance regressors.

Please cite this article in press as: Thomas et al., Predictions and errors are distinctly represented across V1 layers, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.04.036

In total, this GLM resulted in 16 beta images (3x ExpCW, ExpCCW, 1x UnexpCW, UnexpCCW, twice in each scanning run). The beta images from this GLM were grouped according to our main experimental conditions such that we repeated the decoding procedure separately four times (Expected [x3], Unexpected) to determine whether stimulus orientation classification differed across each of these conditions. We tested each of these four decoding iterations separately, restricting the voxels to each of our three V1 layer masks. This procedure resulted in 12 testing iterations (4 conditions in each of 3 masks). Accuracy of the SVM was calculated as the proportion of correctly classified images across all decoding steps and was conducted separately for each participant. The accuracy scores for each of the three expected conditions were averaged for each participant, providing one accuracy score for 'expected' trials, and one for 'unexpected' trials, in each of the layer masks. These scores were then compared between expected and unexpected conditions, and across layers, to determine whether information about presented stimuli varied as a function of learned expectation across the cortical layer bins.

The results were then analysed with a 2x3 repeated measures ANOVA with the factors experimental condition (expected, unexpected) and cortical layer (deep, middle, superficial). Follow up tests examined differences across layers, separately for expected and unexpected events.