Reviewed Preprint
v1 • May 8, 2026
Not revised

Reviewed Preprint
v2 • August 27, 2026
Revised by authors

A Functional Influence Based Circuit Motif That Constrains the Set of Plausible Algorithms of Cortical Function

Anna Vasilevskaya[1,2], **Georg B Keller**[1,2]

[1]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland • [2]Faculty of Science, University of Basel, Basel, Switzerland

For correspondence: georg.keller@fmi.ch

Competing interests: No competing interests declared

Funding: See page 36

Reviewing editor: Rui Ponte Costa, University of Oxford, United Kingdom

eLife Assessment

This **fundamental** work significantly advances our understanding of the circuit-level implementation of predictive processing by elucidating the functional influence between putative prediction error neurons in layer 2/3 and putative internal representation neurons in layer 5. The evidence demonstrating that neither the hierarchical nor the non-hierarchical variant of predictive processing fully accounts for the presented data is **convincing**. Moving forward, this line of work would benefit from explicitly comparing different theories, thereby clearly articulating the points raised in this paper.

https://doi.org/10.7554/eLife.110827.2.sa4

Abstract

There are several plausible algorithms for cortical function that are specific enough to make testable predictions of the interactions between functionally identified cell types. Many of these algorithms are based on some variant of predictive processing. Here we set out to experimentally distinguish between two such predictive processing variants. A central point of variability between them lies in the proposed vertical communication between layer 2/3 and layer 5, which stems from the diverging assumptions about the computational role of layer 5. One assumes a hierarchically organized architecture and proposes that, within a given node of the network, layer 5 conveys unexplained bottom-up input to prediction error neurons of layer 2/3. The other proposes a non-hierarchical architecture in which internal representation neurons of layer 5 provide predictions for the local prediction error neurons of layer 2/3. We show that the functional influence of layer 2/3 cell types on layer 5 is incompatible with the hierarchical variant, while the functional influence of layer 5 cell types on prediction error neurons of layer 2/3 is incompatible with the non-hierarchical variant. Given these data, we can constrain the space of plausible algorithms of cortical function. We propose a model for cortical function based on a combination of a joint embedding predictive architecture (JEPA) and predictive processing that makes experimentally testable predictions.

Introduction

The fact that the layered organization of neocortex is largely consistent across areas with very different function has been a key driver of the idea that all cortical areas are governed by a general algorithm (Mumford, 1992). A search for candidates for such an algorithm should be constrained to those that perform a useful computation when implemented *in silico*, generalize to all known functions of cortex, and can be mapped onto cortical cell types. The most promising proposals for such an algorithm have been different variants of predictive processing. At its core, predictive processing is based on the idea that cortex functions by learning and maintaining

internal models of the world that shape predictions about the upcoming signals from the world (Mumford, 1992; Rao and Ballard, 1999). In predictive processing, these signals are typically referred to as sensory signals. We will refer to them as ***teaching signals***. The terminology of sensory signals is inadequate in non-hierarchical networks or away from the sensory periphery, where the world consists merely of other brain regions, and signals do not necessarily correspond to any specific sensory modality. Thus, following Jordan and Rumelhart (Jordan and Rumelhart, 1992) we will use the more general term of teaching signal to refer to this type of signal. We can formally define a teaching signal as a *signal that functionally serves as a reference – or a target – for a given prediction signal* (Aizenbud et al., 2025). Teaching signals are compared with their corresponding predictions, and the resulting prediction errors update the internal representation and drive plasticity to update the internal model. In its most common variant predictive processing postulates that prediction errors are computed in layer 2/3 neurons (Figure 1A) (Bastos et al., 2012; Keller and Mrsic-Flogel, 2018). Due to low baseline firing rates in layer 2/3, two classes of prediction error neurons were postulated – positive prediction error neurons that signal when the teaching signal exceeds the prediction, and negative prediction error neurons that signal when the prediction exceeds the teaching signal. These two classes of prediction error neurons can be mapped onto two different molecularly defined cell types of layer 2/3 (O'Toole et al., 2023). Positive and negative prediction errors exert opposing influence on internal representation neurons, a mechanism that facilitates the bidirectional integration of error signals. Internal representation neurons thus represent the estimate of the corresponding teaching signal. They are also the source of prediction and teaching signals sent to other cortical areas. Based primarily on anatomical considerations, it is often assumed that internal representation neurons are a subset of layer 5 excitatory neurons (Bastos et al., 2012; Keller and Mrsic-Flogel, 2018).

How prediction error neurons and representation neurons interact, however, differs in different variants of predictive processing (Spratling, 2017). The first proposal was a hierarchical variant of predictive processing in which internal representation neurons act as a source of the teaching signal for the local prediction error neurons and the next higher level of the hierarchy, while simultaneously providing a prediction signal for the preceding lower level of the hierarchy (Figure 1B) (Rao and Ballard, 1999). A later proposal was based on the idea of allowing for non-hierarchical arrangements of networks (Keller and Mrsic-Flogel, 2018). In this variant, the internal representation neuron interacts with the local prediction error neurons of layer 2/3 like a prediction signal (Figure 1C). The idea behind the non-hierarchical implementation was two-fold. First, each node of the network should be able to work independently. With the internal representation acting as a local prediction, a network with just one node will function like a simple temporal difference detector. Second, the exchange of teaching signals and predictions should be possible between any two nodes in a network. Predictions exchanged between primary sensory areas (Garner and Keller, 2022), are likely an example of a non-hierarchical interaction where predictions and teaching signals are exchanged bidirectionally. The non-hierarchical variant of predictive processing anticipates that error neurons can integrate over multiple sources of predictions. It should be noted that neither model, of course, represents the only possible solution for hierarchical/non-hierarchical organization of cortical nodes, we just use this terminology to name the two models based on their most important conceptual difference. As a result of the differences in implementation, the two models make experimentally distinguishable predictions about the interactions between representation neurons and prediction error neurons. Essential is the fact that all proposed interactions between prediction error neurons and local internal representation neurons have reversed sign in the two models (Figures 1B and 1C).

In this work we focus on testing these predictions in mouse primary visual cortex (V1) in the context of visuomotor interactions. Our approach was based on the ability to functionally and molecularly identify layer 2/3 positive and negative prediction error neurons and probe their interaction with other genetically identified neuron types in cortex using optogenetics. If the hierarchical model is correct, we expect to find a genetically identified neuron type (the still unidentified internal representation neuron) in the local circuit that excites positive prediction error neurons and inhibits negative prediction error neurons, while at the same time we expect this representation neuron type to be inhibited by positive prediction error neurons and excited

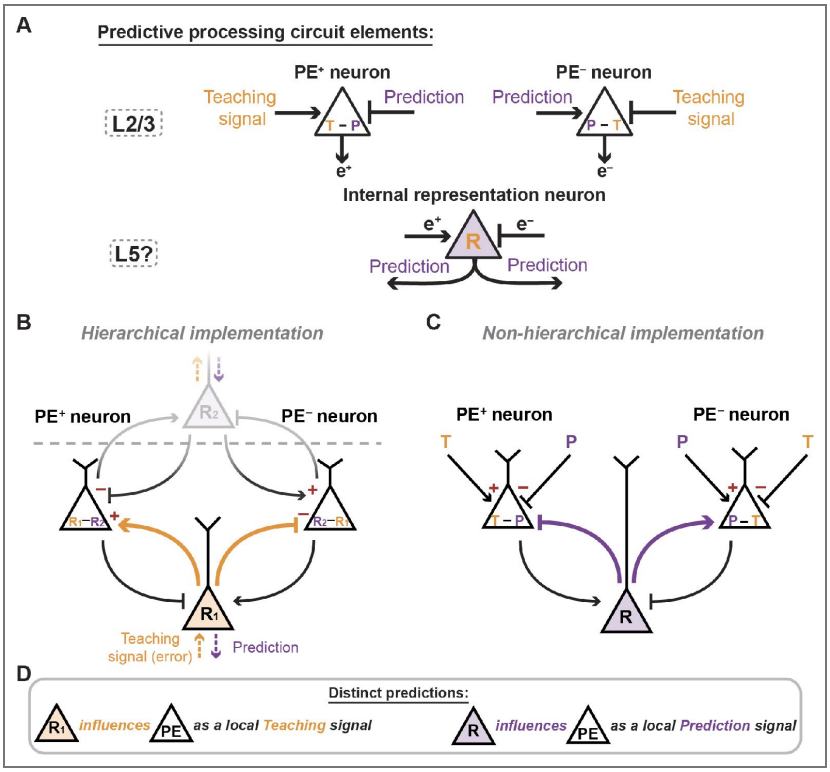


Figure 1. Predictive processing circuit model variants postulating a testable motif of interactions between neurons of deep and superficial cortical layers.

(A) Two classes of neurons postulated by predictive processing. Top: A comparator circuit. Prediction error neurons compute the difference between the teaching signal and prediction. Positive prediction errors (**PE+**) are computed as Teaching signal minus Prediction (T – P), negative prediction errors (**PE-**) are computed as Prediction minus Teaching signal (P – T). Here and in the subsequent panels, arrows represent excitatory connections and blunt arrows inhibitory connections (inhibitory interneurons are omitted for simplicity – see (Attinger et al., 2017; Widmer et al., 2022) for a full circuit description). Bottom: Internal representation circuit. Internal representation neurons (**R**) integrate prediction errors and shape the prediction (or teaching) signals sent to other areas. **(B)** Schematic of the hierarchical implementation of predictive processing. **(C)** Schematic of the non-hierarchical implementation of predictive processing. **(D)** The distinct predictions made by hierarchical and non-hierarchical implementations on the functional influence of deep layer neurons on superficial layer neurons.

by negative prediction error neurons (Figure 1B). If the non-hierarchical model is correct, we expect to find a genetically identified neuron type in the local circuit that inhibits positive prediction error neurons and excites negative prediction error neurons, again with reversed influence in the opposite direction (Figure 1C).

Results

In an initial set of experiments, we aimed to test the predictions made by the two circuit models (Figure 1) regarding the functional influence that internal representation neurons exert on local prediction error neurons. Critical for the interpretation of our results here is the assumption that internal representation neurons are a genetically identifiable cell type in layers 5 or 6. Thus, we screened the functional influence of genetically identified subsets of excitatory neurons in layers 5 and 6 on functionally identified prediction error neurons in layer 2/3. To be able to record the activity of layer 2/3 neurons, we expressed a genetically encoded calcium indicator by injecting an AAV2/1-EF1α-GCaMP6f vector in visual cortex. We did this in three different mouse lines that express Cre recombinase in a subset of layer 5 intratelencephalic (IT, Tlx3-Cre), layer 5 extratelencephalic (ET, Fezf2-Cre) or layer 6 (Ntsr1-Cre) neurons. To be able to optogenetically activate these neurons, we also injected an AAV2/1-hSyn-DIO-ChrimsonR-tdTomato vector to express the optogenetic activator ChrimsonR in Cre positive neurons in primary visual cortex (Figure 2A). For all experiments, mice were head fixed on a spherical treadmill surrounded by a toroidal screen that displayed a virtual reality environment. The virtual environment consisted of a tunnel, with walls patterned with vertical sinusoidal gratings (Figure 2B). We used a two-photon microscope with a coaxial 637 nm stimulation laser to activate ChrimsonR while recording calcium responses in layer 2/3 neurons (**Methods**).

To functionally identify prediction error neurons in layer 2/3 of V1, we used the virtual reality system to present a set of visuomotor stimuli that are known to elicit different types of prediction error responses in mouse V1 (Jordan and Keller, 2020; O'Toole et al., 2023). Mice were exposed first to a closed loop condition, in which visual flow feedback in the virtual tunnel was coupled to locomotion on a spherical treadmill. In the closed loop condition, we presented brief (1 s) unpredictable halts in visual flow, thus briefly breaking the coupling between locomotion and visual flow feedback. We refer to such a break in coupling as a visuomotor mismatch. The responses to a visuomotor mismatch can be interpreted as signaling a negative prediction error (Keller and Mrsic-Flogel, 2018; Keller and Sterzer, 2024; Rao and Ballard, 1999), since during a visuomotor mismatch there is less visual flow than predicted given locomotion (Figure 2C, bottom). After the closed loop condition, mice were exposed to an open loop condition. In the open loop condition, locomotion and visual flow feedback were uncoupled and mice viewed a replay of the visual flow that they had self-generated in the preceding closed loop session. Finally, during a grating session, full-field gratings of randomized orientation, direction, and duration were presented to the mouse at random times (**Methods**). In both open loop and grating sessions, the visual stimuli were presented independently of the locomotion behavior of the mouse.

The presentation of unpredictable grating stimuli results in increases of activity in a subset of layer 2/3 neurons that are typically referred to as visual, or sensory, responses. As these stimuli constitute more visual input than predicted by movement (Figure 2C, top), such responses can be interpreted as positive prediction error responses (Keller and Mrsic-Flogel, 2018; Keller and Sterzer, 2024; Rao and Ballard, 1999). Note, in predictive processing multiple neuron types are expected to have sensory driven increases in activity, positive prediction error neurons are just one of these (Keller and Mrsic-Flogel, 2018). As a note on terminology, we will refer to 'visuomotor mismatch' and 'grating onset' responses, after the stimuli used to trigger them, and to 'negative and positive prediction error neurons', when talking about the neurons in layer 2/3 that are responsive to these stimuli.

Both visuomotor mismatch and grating onset elicited strong population responses in layer 2/3. For each stimulus, the distribution of individual neuronal responses was composed of a continuum, ranging from strong increases in activity to decreases in activity (Figures 2D and 2E). We selected the 15% of neurons most responsive to grating onset and will refer to these as positive

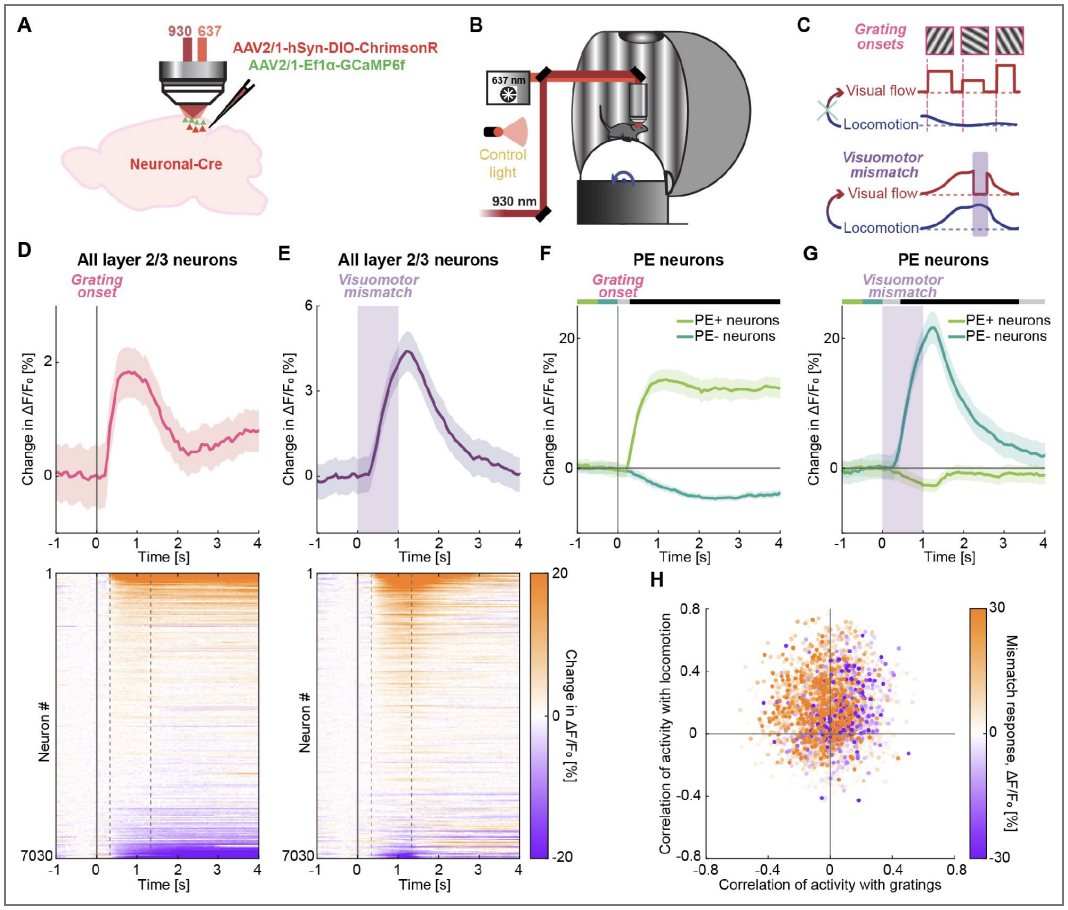


Figure 2. Functional identification of positive and negative prediction error neurons in layer 2/3 of V1.

(A) AAV vector injections were used to express ChrimsonR-tdTomato in different Cre-positive deep layer neurons (Figures 3-5), and GCaMP6f in layer 2/3 neurons of V1. The ChrimsonR virus was omitted in no-ChrimsonR control mice. **(B)** Schematic of the two-photon microscope and a virtual reality system. **(C)** Schematic of the visuomotor stimuli used to probe for prediction error responses. Grating onsets were used to identify positive prediction error neurons and visuomotor mismatch stimuli were used to identify negative prediction error neurons. **(D)** Top: Mean layer 2/3 population response to grating onset. Here and in the subsequent panels, the solid line represents the hierarchical bootstrap estimates of the mean trace. Shading around the mean is the bootstrap error, defined as one standard deviation of the bootstrap distribution at each time bin. Bottom: Heatmap of mean responses of all layer 2/3 neurons. Neurons are sorted by the amplitude of their grating onset response, dashed gray lines mark the sorting window. **(E)** As in **D**, but for visuomotor mismatch. Purple shading marks the duration of the mismatch stimulus. **(F)** Mean response of positive (light green) and negative (dark green) prediction error neurons to grating onset. Here and in the subsequent panels, response curves are compared for each time bin: in the horizontal bars above the plot, black marks time bins where $p<0.05$ and gray mark time bins where $p>0.05$. Colored bars to the left indicate which lines are compared. **(G)** Mean response of positive (light green) and negative (dark green) prediction error neurons to visuomotor mismatch. **(H)** Scatter plot of the correlation between neuronal activity and optical flow speed during grating sessions against the correlation between neuronal activity and locomotion speed for all layer 2/3 neurons. Dot color indicates the amplitude of the response to visuomotor mismatch.

prediction error neurons. By construction, positive prediction error neurons exhibited larger grating onset responses than the population average (Figures 2D and 2F) but exhibited negative responses to visuomotor mismatch (Figure 2G). Similarly, we selected the 15% most responsive neurons to visuomotor mismatch and will refer to these as negative prediction error neurons. Note, in both cases our results do not depend on the exact percentage of neurons included here (Figures S3, S6, S7, **and** S9). Consistent with previous findings (Attinger et al., 2017), negative prediction error neurons exhibited a reduction in response to the presentation of full-field gratings (Figure 2F). Also as previously reported (Widmer et al., 2022), their responses to locomotion onset depended on whether the onset occurred in closed or open loop conditions (Figure S1). Locomotion onset responses of layer 2/3 were on average stronger in the open loop condition (including grating sessions) than in the closed loop condition (Figure S1A), possibly through a visual flow induced inhibition – since a closed loop locomotion onset is always accompanied by a visual flow onset. Locomotion onset responses of positive prediction error neurons were smaller than those of negative prediction error neurons and were stronger in closed loop than in open loop (Figure S1B). This response profile is also consistent with the functional and computational features of positive prediction error neurons that are thought to compute the difference between visual teaching input and motor related prediction (Jordan and Keller, 2020; Keller and Mrsic-Flogel, 2018). Finally, prediction error neurons in layer 2/3 exhibit differential correlations with visual flow and locomotion speed in open loop conditions (Attinger et al., 2017). Plotting the distribution of layer 2/3 neuronal activity correlations with a binarized grating trace against the activity correlations with locomotion speed shows that mismatch responsive neurons preferentially exhibit negative correlations with visual input and positive correlations with locomotion (Figure 2H). An opposing influence of a negative visual teaching input and a positive motor related prediction signal is a hallmark feature of negative prediction error neurons in V1 (Jordan and Keller, 2020).

Thus, we can separate layer 2/3 neurons into putative positive and negative prediction error neurons based on their functional responses. To test the functional influence of layer 5 IT neurons on these subpopulations of layer 2/3 neurons, we stimulated the Tlx3 neurons optogenetically at random times while recording layer 2/3 responses (Figure 3A). We found that we could elicit responses in layer 2/3 neurons when stimulating layer 5 IT neurons (Figure 3B). To control for responses that were elicited independently of the optogenetic activation, we used both a control light outside of the craniotomy, to emulate purely visual responses in the same mice, and a 'no-ChrimsonR' control using the stimulation light in a different cohort of mice, in which we omitted the injection of the ChrimsonR vector. In both control conditions, we found visual responses to the control or the stimulation light turning on (Figure 3B). Thus, even though the stimulation light is red (637 nm), and outside the peak sensitivity of the mouse retina, it appears to be bright enough to be visible to the mice. However, these responses were considerably smaller than the optogenetic stimulation responses (Figure 3B) and activated a different set of neurons. Sorting neurons by their response strength to optogenetic stimulation of layer 5 IT neurons revealed a spectrum of responses (Figure 3C). Responses of the same neurons to control light stimulation were only poorly correlated with the responses to layer 5 IT stimulation (Figures 3C and S2A). In the no-ChrimsonR control mice, the response distribution was similar to that in the control light stimulation case (Figure S2B). Based on this, we conclude that the optogenetic activation of layer 5 IT neurons triggers responses in layer 2/3 neurons. Note, while there are several indirect pathways from layer 5 IT to layer 2/3 that could contribute to this influence, we see no reason to assume that the majority of the influence is not via direct projections from layer 5 to layer 2/3 (Hage et al., 2022; MICrONS Consortium, 2025).

To test predictions of the two circuit model types (Figure 1) and determine whether the functional influence of layer 5 on layer 2/3 depends on the functional response profile of the layer 2/3 neuron, we analyzed the stimulation responses of positive and negative prediction error neurons (as defined in Figure 2) separately. We found that positive prediction error neurons increased their activity on average, while negative prediction error neurons decreased their activity in response to layer 5 IT stimulation (Figures 3D and 3E). Again, we plotted the correlation of activity with gratings against that with locomotion for each neuron, and color coded

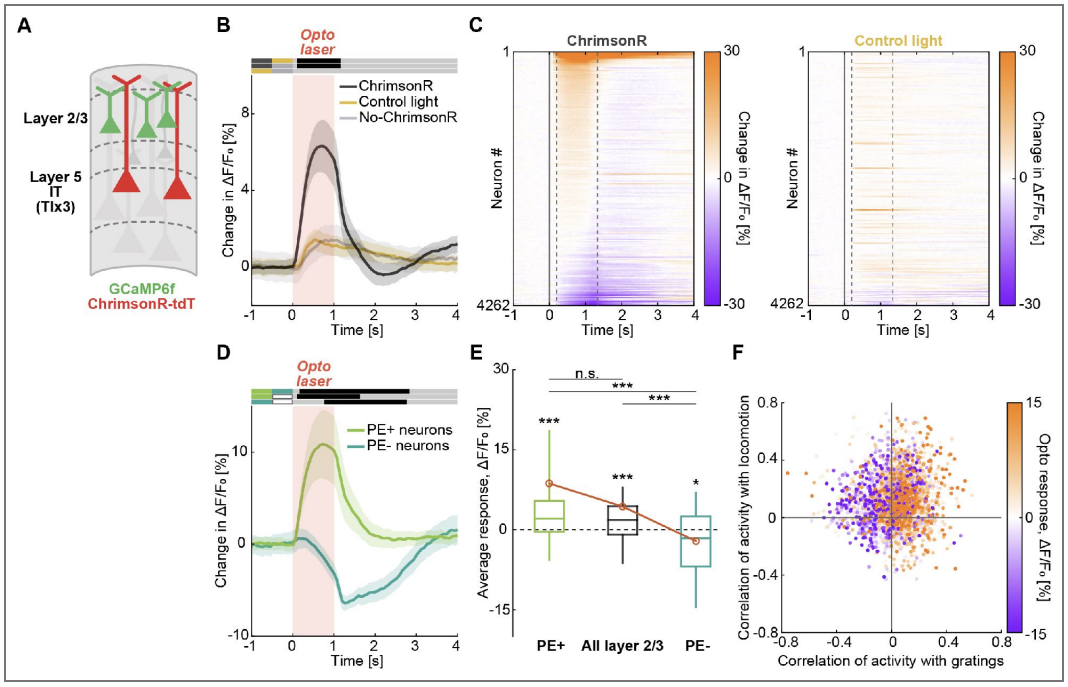


Figure 3. Opposing functional influence of Tlx3 layer 5 neurons on positive and negative prediction error neurons of layer 2/3.

(A) We optogenetically activated Tlx3 layer 5 neurons while recording calcium activity in layer 2/3 neurons. **(B)** Mean layer 2/3 population responses to optogenetic stimulation of Tlx3 layer 5 neurons (dark gray), control light stimulation in the same mice (yellow), and stimulation in no-ChrimsonR control mice (light gray). Red shading marks the stimulation interval. **(C)** Heatmap of mean layer 2/3 neuronal responses to optogenetic stimulation of Tlx3 layer 5 neurons (left) and control light stimulation (right). Neurons are sorted by the amplitude of their response to the optogenetic stimulation. **(D)** Mean responses of positive (light green) and negative (dark green) prediction error neurons in layer 2/3 to optogenetic stimulation of Tlx3 layer 5 neurons. Red shading marks the optogenetic stimulation interval. **(E)** Box plots showing average responses of positive prediction error neurons (light green), negative prediction error neurons (dark green), and all layer 2/3 neurons (black) to optogenetic stimulation of Tlx3 layer 5 neurons. Boxes indicate the interquartile range, the central line marks the median, and whiskers the 10th and 90th percentiles. Orange circles mark the mean response. Here and elsewhere, n.s.: not significant, *: p<0.05, **: p<0.01, ***: p<0.001. **(F)** Scatter plot of the correlation between neuronal activity and optical flow speed against the correlation between neuronal activity and locomotion speed during grating sessions for all layer 2/3 neurons. Dot color indicates the amplitude of the response to optogenetic stimulation of Tlx3 layer 5 neurons.

the response strength to optogenetic stimulation of layer 5 IT neurons (Figure 3F). Similar to the pattern of mismatch responses, we found that neurons that decrease activity on stimulation of layer 5 IT neurons preferentially exhibited positive correlation with locomotion and negative correlation with gratings. This functionally specific modulation of layer 2/3 neurons was preserved if we used different fractions of mismatch and grating onset responsive neurons (Figures S3A-S3C). In contrast, no such modulation was observed with control light stimulation (Figure S3D) or in no-ChrimsonR control mice (Figure S4). Assuming layer 5 IT neurons are internal representation neurons, this would be exactly the relationship postulated by the hierarchical variant of predictive processing (Figure 1B), and the opposite of influence postulated by the non-hierarchical variant (Figure 1C). Thus, we can conclude that the influence of layer 5 IT neurons on layer 2/3 is consistent with the hierarchical variant of predictive processing under the assumption that layer 5 IT neurons are internal representation neurons, but inconsistent with the non-hierarchical variant.

The stimuli we used to identify positive and negative prediction error neurons, visuomotor mismatch and grating onset, only activate a subset of all prediction error neurons. Both sensory and visuomotor mismatch responses (Zmarz and Keller, 2016) exhibit tuning, such that an individual neuron will only respond to a subset of all possible stimuli. While it is not experimentally feasible to cover the full stimulus space, we began to address whether the opposing influence of layer 5 IT stimulation generalizes to other visual and mismatch stimuli. To this end, we repeated the analysis using a limited alternative set of visual and mismatch stimuli. In the case of visual stimuli, we can take advantage of the fact that layer 2/3 neurons are tuned to grating orientation (Figure S5A). Independent of which grating orientation we use to identify positive prediction error neurons, we find the same opposing influence of layer 5 IT stimulation (Figure S5B). For visuomotor mismatch responses, we performed an additional set of experiments in which we exposed mice both to normal coupling, with forward locomotion resulting in backward visual flow, and to an inverted coupling, where forward locomotion resulted in forward visual flow (Figure S5C). In both conditions we measured visuomotor mismatch responses in layer 2/3 neurons. We found that the two types of visuomotor mismatch tended to activate different sets of neurons (Figure S5D). Like normal visuomotor mismatch neurons, inverted visuomotor mismatch neurons also exhibited a decrease in activity upon visual stimulation (Figure S5E). Importantly, we found that inverted visuomotor mismatch neurons also exhibited the same sign of layer 5 IT influence we had seen for normal visuomotor mismatch neurons (Figure S5F). Thus, we speculate that layer 5 IT neurons generally have an excitatory functional influence on positive prediction error neurons and an inhibitory functional influence on negative prediction error neurons in layer 2/3.

Our interpretation of the results relies on an assumption of which cell type functions as internal representation neuron. It is possible that a different cell type in layers 5 or 6 might exert an inverted pattern of influence on layer 2/3. Thus, we repeated our experiments with a Cre line that labels layer 5 ET neurons (Fezf2-Cre) and one that labels layer 6 neurons (Ntsr1-Cre). Interestingly, we found the same opposing influence of these neuron types on positive and negative prediction error neurons in layer 2/3. Positive prediction error neurons increased their activity upon stimulation of both layer 5 ET (Figures 4 and S6) and layer 6 neurons (Figures 5 and S7), while negative prediction error neurons decreased their activity on stimulation – very similar to what we found for layer 5 IT neurons (Figure 3). Thus, we found no evidence of an interaction between deep cortical layers and layer 2/3 consistent with internal representation neurons and the non-hierarchical variant of predictive processing (Figure 1C), but rather all interactions are consistent with internal representation neurons and the hierarchical variant of predictive processing (Figure 1B). Hence, quite remarkably, we find that the functional response profile of layer 2/3 neurons is predictive of the influence that they receive from stimulation of deep cortical layers. Moreover, this influence is consistent with that predicted by the hierarchical predictive processing model.

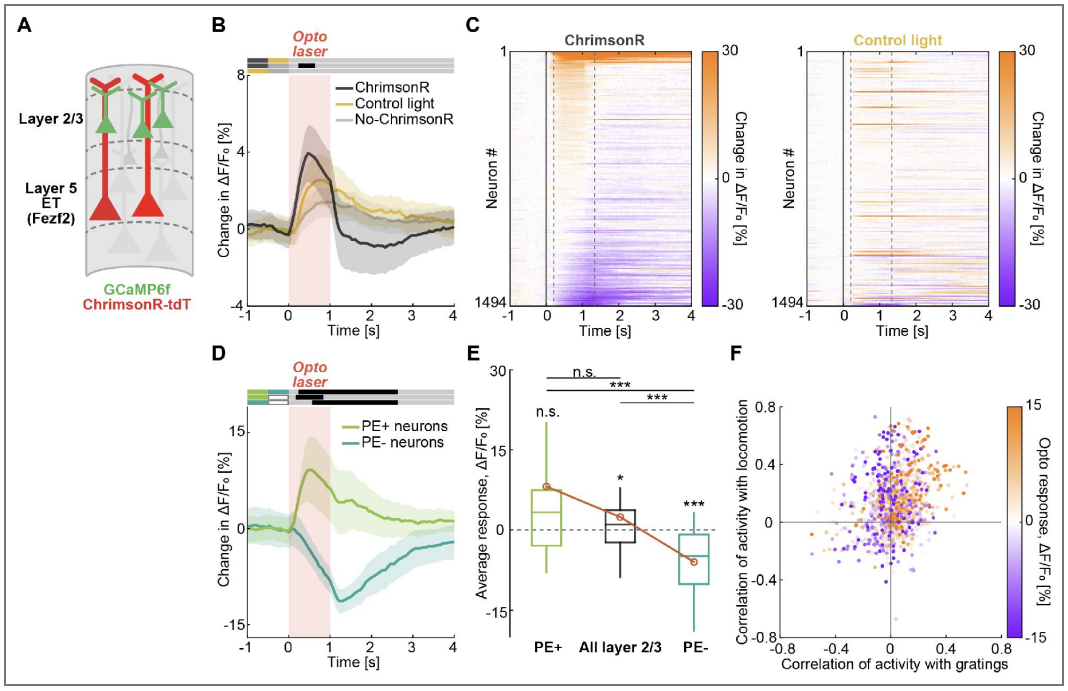


Figure 4. Opposing functional influence of Fezf2 layer 5 neurons on positive and negative prediction error neurons of layer 2/3.

(A) We optogenetically activated Fezf2 layer 5 neurons while recording calcium activity in layer 2/3 neurons. **(B)** Mean layer 2/3 population responses to optogenetic stimulation of Fezf2 layer 5 neurons (dark gray), control light stimulation in the same mice (yellow), and stimulation in no-ChrimsonR control mice (light gray). Red shading marks the stimulation interval. **(C)** Heatmap of mean layer 2/3 neuronal responses to optogenetic stimulation of Fezf2 layer 5 neurons (left) and control light stimulation (right). Neurons are sorted by the amplitude of their response to the optogenetic stimulation. **(D)** Mean responses of positive (light green) and negative (dark green) prediction error neurons in layer 2/3 to optogenetic stimulation of Fezf2 layer 5 neurons. Red shading marks the optogenetic stimulation interval. **(E)** Box plots showing average responses of positive prediction error neurons (light green), negative prediction error neurons (dark green), and all layer 2/3 neurons (black) to optogenetic stimulation of Fezf2 layer 5 neurons. Boxes indicate the interquartile range, the central line marks the median, and whiskers the 10th and 90th percentiles. Orange circles mark the mean response. **(F)** Scatter plot of the correlation between neuronal activity and optical flow speed against the correlation between neuronal activity and locomotion speed during grating sessions for all layer 2/3 neurons. Dot color indicates the amplitude of the response to optogenetic stimulation of Fezf2 layer 5 neurons.

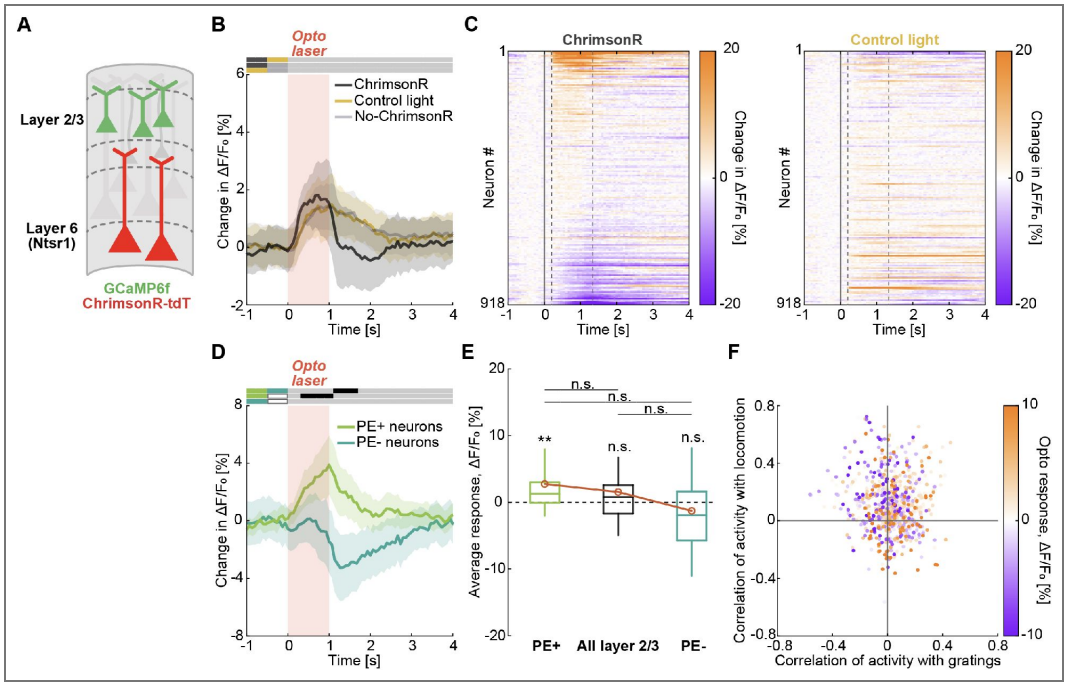


Figure 5. Functional influence of Ntsr1 layer 6 neurons on positive and negative prediction error neurons of layer 2/3.

(A) We optogenetically activated Ntsr1 layer 6 neurons while recording calcium activity in layer 2/3 neurons. **(B)** Mean layer 2/3 population responses to optogenetic stimulation of Ntsr1 layer 6 neurons (dark gray), control light stimulation in the same mice (yellow), and stimulation in no-ChrimsonR control mice (light gray). Red shading marks the stimulation interval. **(C)** Heatmap of mean layer 2/3 neuronal responses to optogenetic stimulation of Ntsr1 layer 6 neurons (left) and control light stimulation (right). Neurons are sorted by the amplitude of their response to the optogenetic stimulation. **(D)** Mean responses of positive (light green) and negative (dark green) prediction error neurons in layer 2/3 to optogenetic stimulation of Ntsr1 layer 6 neurons. Red shading marks the optogenetic stimulation interval. **(E)** Box plots showing average responses of positive prediction error neurons (light green), negative prediction error neurons (dark green), and all layer 2/3 neurons (black) to optogenetic stimulation of Ntsr1 layer 6 neurons. Boxes indicate the interquartile range, the central line marks the median, and whiskers the 10th and 90th percentiles. Orange circles mark the mean response. **(F)** Scatter plot of the correlation between neuronal activity and optical flow speed against the correlation between neuronal activity and locomotion speed during grating sessions for all layer 2/3 neurons. Dot color indicates the amplitude of the response to optogenetic stimulation of Ntsr1 layer 6 neurons.

To further test these models, we investigated the predictions they make about the influence in the reverse direction, from prediction error neurons on internal representation neurons (Figure 6A). This is technically a bit more challenging as it requires a way to target optogenetic stimulation specifically either to positive or to negative prediction error neurons in layer 2/3. To do this, we used two different artificial regulatory elements that can be used in AAV vectors to bias expression of a reporter to either positive or negative prediction error neurons (O'Toole et al., 2023) (Figure 6B). The AP.Baz1a.1 regulatory element biases expression to positive prediction error neurons (the Rrad transcriptional cell type), and the AP.Adamts2.1 regulatory element biases expression to negative prediction error neurons (the Adamts2 transcriptional cell type). We will refer to neurons labeled using this strategy by the target cell type of the corresponding artificial regulatory element (Rrad and Adamts2). We performed these experiments in two separate cohorts of mice. In the first, we expressed ChrimsonR under the control of AP.Baz1a.1 regulatory element and recorded calcium responses in Tlx3 layer 5 IT neurons (Figure 6C). We found that optogenetic activation of Rrad neurons resulted in a trend towards increased activity in layer 5 IT neurons (Figures 6D and 6G). Repeating the experiments in mice that expressed ChrimsonR under the control of the AP.Adamts2.1 regulatory element (Figure 6E), we found that optogenetic activation of Adamts2 neurons resulted in a net decrease of activity in layer 5 IT neurons (Figures 6F and 6G). This interaction is the inverse of what the hierarchical version of predictive processing would postulate under the assumption that layer 5 neurons function as internal representation neurons, but matches the predictions made by the non-hierarchical variant under the same assumption (Figure 6A).

Where does that leave us? Neither variant of the model can account for the influence between superficial and deep layers under the assumption that prediction error neurons are in layer 2/3 and internal representation neurons in layers 5 or 6. There are a number of caveats to this conclusion that originate in the fact that our approach may not have covered all possible subtypes of deep cortical neurons (see **Discussion**). But let us call this a lemma for now: either both predictive processing models are wrong, or internal representation neurons are not in deep cortical layers.

As a starting point for our argument to explain these observations, it is important to note that from the perspective of prediction error neurons in layer 2/3, the influence of deep cortical layers is consistent with a teaching signal. Based on this, we speculated that the prediction error responses in layer 2/3 may not be computed with respect to sensory input, but with respect to layer 5 activity as a teaching signal. We noticed that the functional influence of layer 5 IT neurons depended on whether the mice were locomoting. The functional influence was decreased while mice were locomoting, but in such a way that there was a population of neurons that decreased their activity stronger during locomotion than during stationary periods (Figures S8A-S8C). These differences could not be explained by differences in the optogenetic activation of the layer 5 IT neurons themselves, since they were similarly activated in both cases (Figures S8D-S8F). Such an unmasking of an inhibitory functional influence would be consistent with a computation in the layer 2/3 negative prediction error neurons that compare an excitatory prediction with an inhibitory teaching signal from layer 5.

Given our results thus far, we hypothesized that the predictive processing network in superficial layers of cortex would function to model and predict activity in deep cortical layers. One way to investigate this idea more directly is to generate artificial activity patterns in layer 5 and test whether layer 2/3 responses look like prediction errors with respect to perturbations of the artificial layer 5 activity patterns. Thus, we designed an experiment in which the optogenetic activation of layer 5 IT neurons was closed loop coupled to the locomotion speed of the mouse without any visual feedback coupling (**Methods**). The idea was to create an artificial closed loop coupling in which layer 2/3 now computes a difference between a motor-related prediction and a teaching signal it receives from the local layer 5 population. To test whether this was the case, we briefly broke the coupling between the optogenetic activation of layer 5 IT neurons and locomotion speed (Figure 7A) at random times for 1 s (akin to how we trigger visuomotor mismatch by halting visual flow) and measured neuronal responses to this perturbation. This

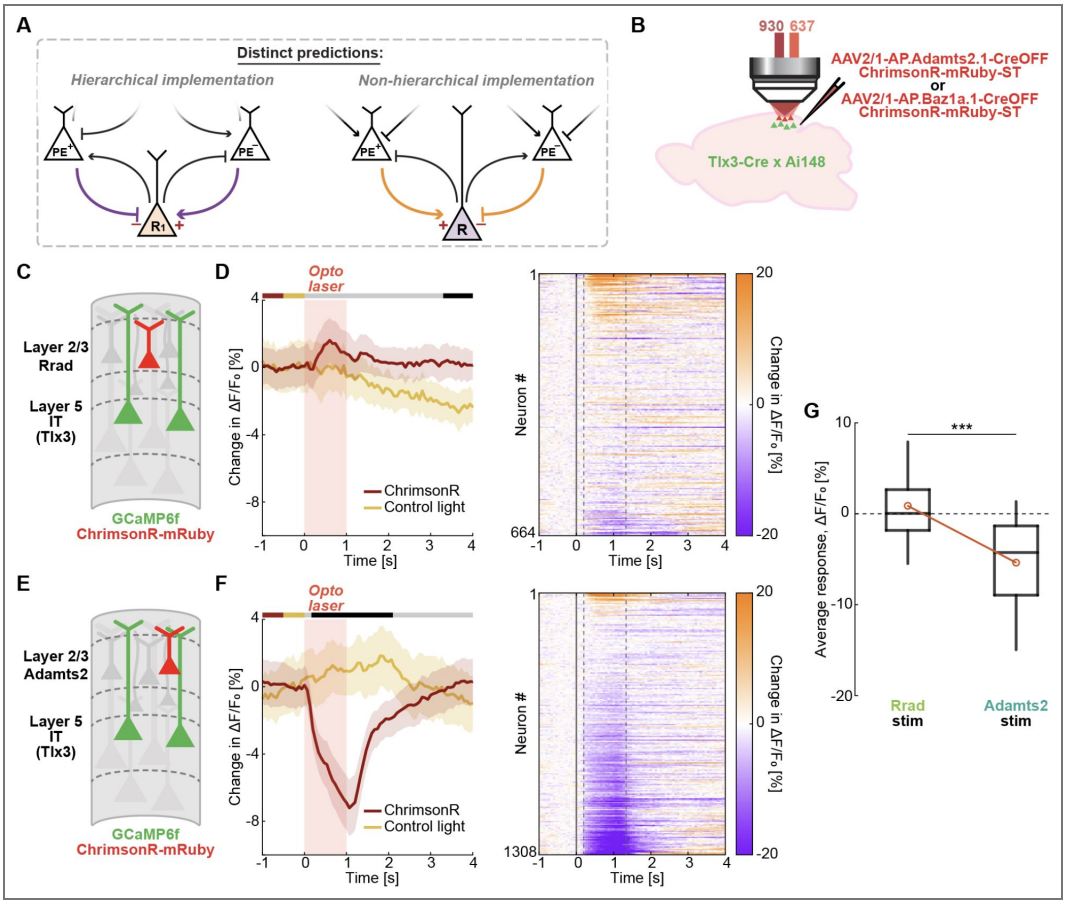


Figure 6. Opposing functional influence of positive and negative prediction error neurons on Tlx3 layer 5 IT neurons.

(A) Schematic of the distinct experimental predictions made by hierarchical and non-hierarchical predictive processing implementations on the functional influence of superficial layer neurons on deep layer neurons. **(B)** Two artificial regulatory elements (either AP.Baz1a.1 or AP.Adamts2.1) were used in AAV vectors to bias expression of soma-targeted (ST) ChrimsonR to either positive (Rrad) or negative (Adamts2) prediction error neurons respectively. AAV vectors were injected in V1 of Tlx3-Cre x Ai148 mice. **(C)** We optogenetically activated Rrad neurons while recording calcium activity in Tlx3 layer 5 IT neurons. **(D)** Left: Mean layer 5 IT populational response to optogenetic stimulation of Rrad neurons (red) and control light stimulation in the same mice (yellow). Red shading marks the stimulation interval. Right: Heatmap of mean layer 5 IT neuronal responses to optogenetic stimulation of Rrad neurons. Neurons are sorted by the amplitude of their response to the optogenetic stimulation. **(E)** We optogenetically activated Adamts2 layer 2/3 neurons while recording calcium activity in Tlx3 layer 5 IT neurons. **(F)** As in **D**, but for stimulation of Adamts2 neurons. **(G)** Comparison of the responses of Tlx3 layer 5 IT neurons to optogenetic stimulation of Rrad or Adamts2 neurons. Boxes indicate the interquartile range, the central line the median, and whiskers the 10th and 90th percentiles. Orange circles mark the mean response.

optomotor mismatch resulted in strong responses in layer 2/3 that could not be explained by the offset of optogenetic stimulation alone (Figure 7B). The optomotor mismatch responses looked surprisingly similar to visuomotor mismatch responses on a population level (Figure 7C). Strikingly, by plotting the responses of all layer 2/3 neurons to optomotor mismatch against their responses to visuomotor mismatch, we found that there is a positive correlation between these two responses across the population, indicating that the same neurons are activated by both visuomotor and optomotor mismatch stimuli (Figure 7D). Thus, these data are consistent with the idea that the visuomotor mismatch responses we observe in layer 2/3 neurons are computed with layer 5 serving as a teaching signal, and that superficial cortical layers function to predict activity in deep cortical layers.

Finally, if the input from deep cortical layers is the teaching signal for the prediction error computation in layer 2/3, what is the role of the visual input to layer 2/3 that arrives via layer 4? To begin addressing this question, we measured the functional influence from Scnn1a layer 4 neurons to layer 2/3 neurons (Figure 8A). Here, we found no evidence that the functional influence depended on whether a given layer 2/3 neuron was a positive or a negative prediction error neuron. (Figures 8B-8F and S9). Compared to functional influence of layer 5 IT, functional influence of layer 4 on layer 2/3 also did not depend on locomotion state of the mouse (Figures S8G-S8I). Thus, Scnn1a layer 4 neurons do not appear to function as a teaching signal for layer 2/3 prediction error neurons. Note, the Scnn1a-Cre line labels only one of the layer 4 cell types (Tasic et al., 2016), and there is direct thalamic input to layer 2/3 (Douglas et al., 1989; Morgenstern et al., 2016). While there may be other layer 4 cell types that function as teaching signal for layer 2/3, our findings indicate that the distinct modulation provided by layers 5 and 6 is not a universal property of all cortical inputs, but is likely less pronounced in circuits classically categorized as bottom-up.

Discussion

Here, we experimentally tested predictions of two circuit models of predictive processing (Figure 1) by mapping the functional influence between neurons of deep and superficial cortical layers. We found that functional influence from layers 5 and 6 on layer 2/3 (Figures 3-5) was consistent with the hierarchical predictive processing model but not with the non-hierarchical variant. Quite peculiarly, we also found that the functional influence in the reverse direction from layer 2/3 on layer 5 (Figure 6) was consistent with the non-hierarchical variant but not with the hierarchical one. Thus, under the assumption that prediction error neurons are located in superficial layers of cortex and internal representation neurons in deep layers, neither variant of the predictive processing model can account for the functional interactions between deep and superficial layers of cortex (Figure S10A).

Limitations of the study

There are at least three caveats to our interpretation of the results:

1. The primary caveat is the partial screening coverage with respect to cell types. The mouse Cre lines and artificial regulatory elements we used to target excitatory neurons in layers 2/3, 4, 5, and 6 likely do not cover the space of all possible neuron types in these layers. In all cases, it is conceivable that there is a cell type we did not target that would have a functional influence profile opposite to what we have measured. In the case of the influence of layer 5 on layer 2/3 neurons, our experiments using layer 5 IT and layer 5 ET lines likely cover the majority of the dominant influence (Kim et al., 2015; Matho et al., 2021). In case of assessing the layer 4 influence on layer 2/3, the Scnn1a-Cre mouse only labels one of the subtypes of layer 4 excitatory neurons (Tasic et al., 2016). Here it is certainly possible that future experiments will find other layer 4 cell types with different profiles of functional influence on layer 2/3. However, our primary conclusions that layer 5 acts like a teaching input to layer 2/3 would not be changed by this. Finally, for the measurements of influence of layer 2/3 on layer 5, we have restricted our experiments to the influence on layer 5 IT neurons. This choice was motivated by the argument that layer 5 IT neurons are likely the primary target of layer 2/3 neurons in layer

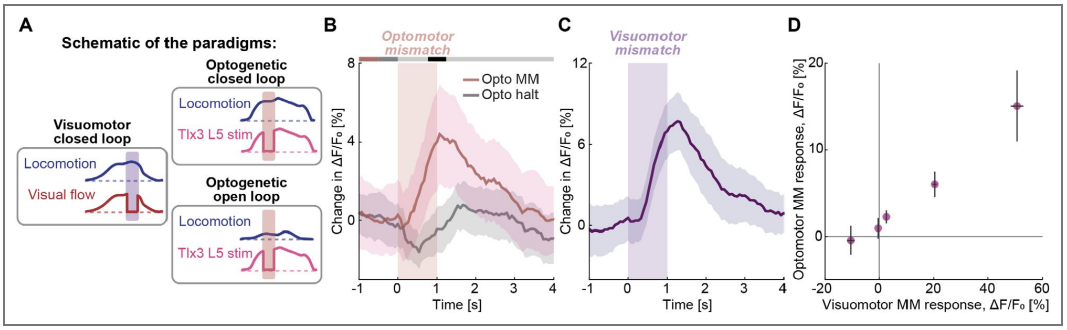


Figure 7. Layer 2/3 responds to manipulations of layer 5 activity as if layer 5 were a teaching signal for layer 2/3.

(A) Mice first experienced a visuomotor closed loop condition, that was then followed by an optogenetic closed loop. In optogenetic closed loop, mice were exposed to constant visual illumination on the screen while the strength of optogenetic stimulation of Tlx3 layer 5 neurons was coupled to locomotion speed of the mouse instead. Following this, mice were exposed to an optogenetic open loop condition, in which the pattern of optogenetic stimulation the mouse had self-generated in the preceding optogenetic closed loop session was replayed. **(B)** Mean layer 2/3 population response to optomotor mismatch and optogenetic open loop stimulation halt. Pink shading indicates the duration of the stimulus. **(C)** Mean layer 2/3 population response to visuomotor mismatch. Purple shading indicates the duration of the visuomotor mismatch stimulus. **(D)** We binned layer 2/3 neurons by their visuomotor mismatch responses into 5 bins and plotted the mean optomotor mismatch responses of the neurons in these bins. Error bars indicate SEM.

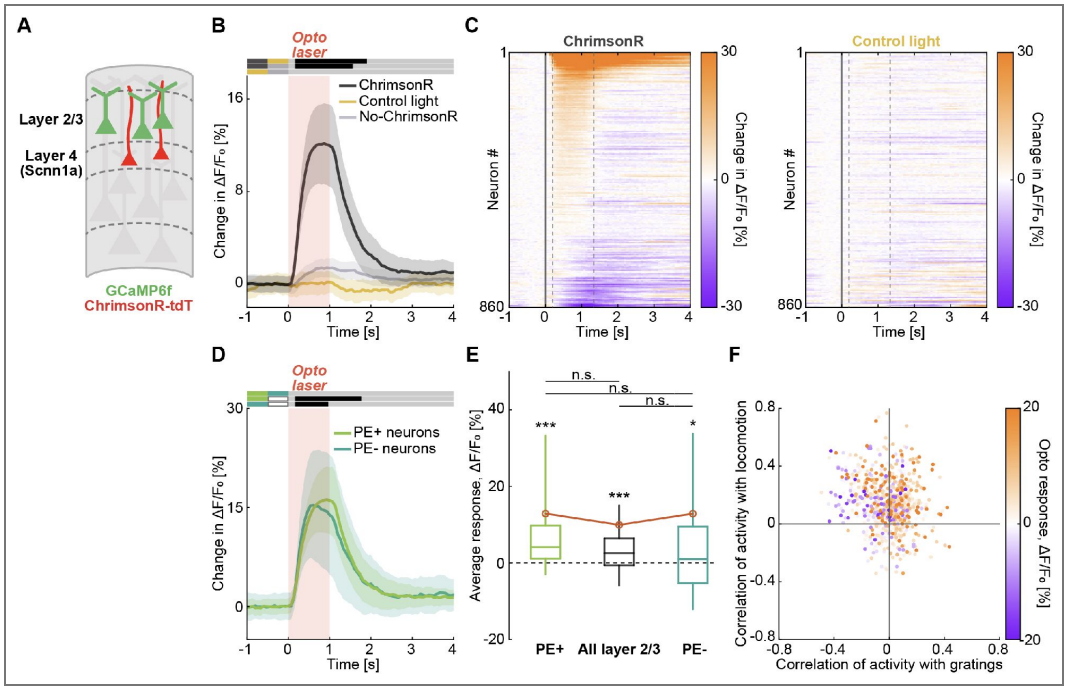


Figure 8. Functional influence of Scnn1a layer 4 neurons on positive and negative prediction error neurons of layer 2/3 is unspecific.

(A) We optogenetically activated Scnn1a layer 4 neurons while recording calcium activity in layer 2/3 neurons. **(B)** Mean layer 2/3 population responses to optogenetic stimulation of Scnn1a layer 4 neurons (dark gray), control light stimulation in the same mice (yellow), and stimulation in no-ChrimsonR control mice (light gray). Red shading marks the stimulation interval. **(C)** Heatmap of mean layer 2/3 neuronal responses to optogenetic stimulation of Scnn1a layer 4 neurons (left) and control light stimulation (right). Neurons are sorted by the amplitude of their response to the optogenetic stimulation. **(D)** Mean responses of positive (light green) and negative (dark green) prediction error neurons in layer 2/3 to optogenetic stimulation of Scnn1a layer 4 neurons. Red shading marks the optogenetic stimulation interval. **(E)** Box plots showing average responses of positive prediction error neurons (light green), negative prediction error neurons (dark green), and all layer 2/3 neurons (black) to optogenetic stimulation of Scnn1a layer 4 neurons. Boxes indicate the interquartile range, the central line marks the median, and whiskers the 10th and 90th percentiles. Orange circles mark the mean response. **(F)** Scatter plot of the correlation between neuronal activity and optical flow speed against the correlation between neuronal activity and locomotion speed during grating sessions for all layer 2/3 neurons. Dot color indicates the amplitude of the response to optogenetic stimulation of Scnn1a layer 4 neurons.

5 (Anderson et al., 2010). Finally, it is likely that the two approaches used for targeting Tlx3 layer 5 neurons in this study (viral labelling versus a cross with a reporter line) may capture only partially overlapping subpopulations.

2. Another major caveat comes from the fact that we bulk stimulate entire populations of neurons. It is possible that each neuron exhibits a center-surround type interaction with its target population, such that a subset of neurons is excited while a larger set of off-target neurons is inhibited. If so, it is possible that in bulk stimulation experiments we would primarily observe the more dominant off-target interactions. If this were the case, all our conclusions on functional influence would be reversed. Assuming this applies to both layer 2/3 influence on layer 5 and vice versa, this would still mean that predictive processing cannot explain our results. It would however mean that layer 5 is not a teaching signal for layer 2/3, but a prediction. However, if our responses were indeed dominated by a surround effect, we would expect to find biphasic responses. Resolving this caveat will require an EM reconstruction approach (Schneider-Mizell et al., 2025).
3. Lastly, we cannot assess what fraction of the stimulation response can be explained by direct versus indirect interactions. For example, in the case of layer 5 stimulation, the influence on layer 2/3 is likely mediated by a combination of direct projections from layer 5 to layer 2/3, as well as via long-range cortical interactions, and subcortical loops. While, based on the relative strength, the reliability of responses, and the projection patterns between layer 5 and layer 2/3 (Hage et al., 2022), we speculate that the direct interaction is dominant, this is not of relevance to our computational interpretations.

Joint embedding predictive architectures: a new starting point for understanding cortex?

There are two algorithms of cortical function we could identify that are consistent with our data. The first is a reciprocal variant of predictive processing, whereas the second is a joint embedding predictive architecture (JEPA). To illustrate the first, there are at least two ways in which predictive processing circuits can be arranged to exchange signals symmetrically. One is the non-hierarchical predictive processing variant we have tested here that is inconsistent with data. In this model both predictions and prediction errors are exchanged bidirectionally. For example, in the communication between visual and auditory cortices, auditory cortex sends a prediction of the visual teaching input to visual cortex given the current auditory representation and vice versa (Garner and Keller, 2022; Keller and Mrsic-Flogel, 2018). Similar arguments have been made for the interactions between sensory and motor areas (Jordan and Rumelhart, 1992; Keller and Sterzer, 2024). As we have established, this arrangement is inconsistent with our data. However, a computationally equivalent implementation would be one in which different cortical areas were to exchange predictions and teaching signals instead of predictions and prediction errors. This requires that each cortical area has two sets of prediction error neurons and two sets of corresponding internal representation neurons. If we add the condition that communication must be symmetrical, this would require two parallel prediction error sub-circuits in each node of the network, with two types of internal representation neurons. For these two sub-circuits the roles of teaching signal and prediction are reversed (Figures S10B-S10C), and so is the connectivity between layer 2/3 and layer 5. Thus, in such a model, we would expect to find both types of opposing influence in both directions in the interactions between layer 2/3 and layer 5. Thus, if we additionally assume that in our experiments, we have only identified parts of each sub-circuit (see caveats), our data would be consistent with this variant of non-hierarchical predictive processing. We speculate that such a reciprocal variant of predictive processing should be implementable given that a restricted hierarchical variant of this idea in the form of bidirectional predictive coding with shared representation neurons has been implemented (Oliviers et al., 2025).

However, we think the reciprocal variant of predictive processing is unlikely the correct model given recent evidence regarding the dominant dimension of coupling in cortex. All variants of predictive processing we discuss here, as well as all other models based on the idea of cortical columns, postulate a dominant role of vertical coupling in cortex: They predict that the activity in

layer 5 of a cortical area is more strongly driven by the activity of layer 2/3 in that same area than by layer 5 activity in surrounding cortical areas. In predictive processing this comes from the assumption of a laminar separation between prediction error neurons and internal representation neurons. Such dominant vertical coupling is hard to reconcile with the fact that psychoactive drugs can selectively alter large scale activity patterns in layer 5 neurons of dorsal cortex without having corresponding effects in layer 2/3 (Bharioke et al., 2022; Heindorf and Keller, 2024; Vesuna et al., 2020). Thus, we conclude that no variant of predictive processing is consistent with both our data and a stronger horizontal than vertical coupling in cortex.

Based on these considerations, we here formulate a new testable hypothesis for the computational algorithm of cortex inspired by the idea of a JEPA (LeCun, 2022). Our proposal centers on the finding that layer 5 functionally interacts with layer 2/3 like a teaching input. Layer 5 input drives layer 2/3 prediction error neurons with a sign consistent with that expected of a teaching signal (Figure 1), and mismatches in the artificial coupling of layer 5 optogenetic activation to locomotion speed are sufficient to drive prediction error responses in layer 2/3 commensurate with those observed in visuomotor coupling (Figure 7). Thus, we propose that layer 2/3 functions to predict layer 5 activity, not sensory input per se, hence making predictions in the internal representation space, not input space. This highlights the key computational difference between JEPA and predictive processing. Predictive processing proposes formation of a generative internal model that constructs predictions in the input space. Consequently, prediction errors are also computed in the input space. In contrast, in a JEPA, predictions are formed in the latent space, and prediction error is computed and minimized in the latent space. According to predictive processing, we would expect the relevant teaching signal for layer 2/3 comparator to be bottom-up input into V1. However, our data instead argues that prediction errors are computed with respect to layer 5 activity as a teaching signal, which is more consistent with a JEPA. Combining this with the fact that deep and superficial layers of cortex each receive thalamic input (Constantinople and Bruno, 2013; Douglas et al., 1989), we speculate that cortex might function like a JEPA (Assran et al., 2023). A JEPA is a deep learning architecture that is designed to learn abstract representations of the input data in a self-supervised manner. It typically consists of two networks where a predictor uses representations of one network to predict the representation of the other (Figure 9A). The loss function in this architecture is based on the comparison between the predicted and the actual target representation. Classic instantiations of this principle in machine learning often use two identical or similar copies of encoder networks that employ some form of weight sharing. In the cortical network, however, it is difficult to conceptualize how such weight-sharing or copying would be implemented. In our analogy to cortex, we speculate that there are two separate networks that symmetrically serve as both encoder and target networks. Specifically, we propose that one of the two networks is housed in the superficial layers whereas the other resides in the deep layers (Figure 9B). The respective predictor networks connecting the two are implemented in columnar connectivity. Superficial and deep cortical layers embed a combination of thalamic and cortical input separately, and layer 2/3 prediction error neurons compute an error between the two embeddings. Thus, a JEPA-like architecture would also predict that the dominant intracortical communication driving activity within deep and superficial layers is horizontal (within layers) and not vertical (across layers), and would thus also simplify the explanation of how deep and superficial layers can be decoupled by psychoactive substances (Keller and Sterzer, 2024).

Beyond JEPA

The core appeal of JEPA lies in its potential to perform a useful computation that can be mapped onto the local cortical microcircuit – a very promising, recently developed JEPA variant with an explicit analogy to cortex is recurrent predictive learning (RPL) (Mohammadi et al., 2025). It proposes two sequential networks, an encoder network whose output is fed to an integrator network. The output of the integrator network is then used to predict the output of the encoder, and the resulting prediction error drives learning. The encoder network would map onto

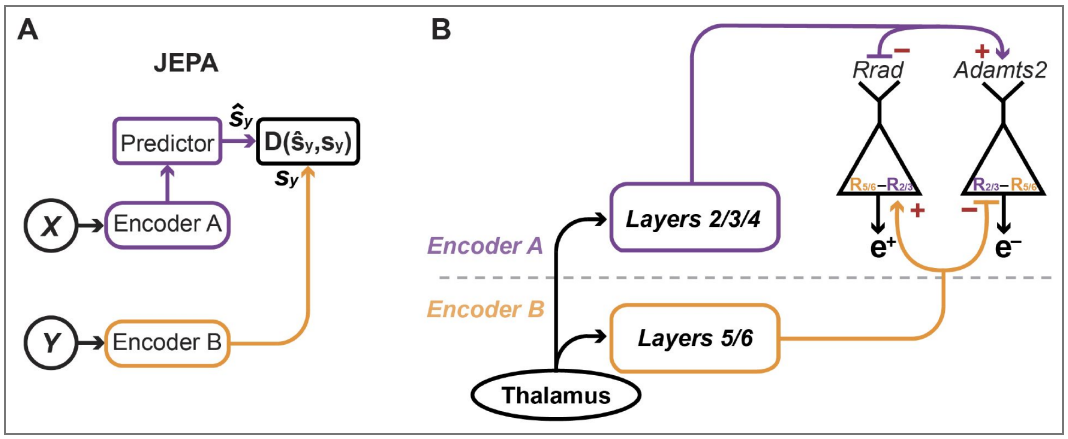


Figure 9. How a JEPA architecture could be mapped onto the cortical circuit.

(A) Schematic of a JEPA architecture for self-supervised learning (adapted from Assran et al., 2023). The output of one encoder A is trained to predict the output of the other encoder B. The prediction is compared to the target output in a comparator network D. **(B)** A proposal for a JEPA implementation in cortical circuits. Superficial and deep cortical layers function as the two separate encoder networks postulated by JEPA. Prediction error neurons in layer 2/3 compare the local layer 5 activity to a prediction of that activity.

superficial layers of cortex and the integrator network onto deep layers. Like a JEPA, RPL would also predict the existence of a prediction error computation that uses the encoder network (deep cortical layers) as a teaching signal.

However, a significant challenge remains: we have not found a way to map the algorithm onto the large-scale cortical architecture. The cortical network, as a whole, is fundamentally non-hierarchical. Most interactions in cortex are cross-modal, associative, or sensorimotor, and not hierarchical. This lack of a clear hierarchical arrangement is also apparent when the principles used to motivate the idea of a cortical hierarchy, based on anatomical data from the visual system (Felleman and Van Essen, 1991), are applied to data collected from across the whole cortex (Markov et al., 2013): Primary somatosensory cortex appears at the top of the "hierarchy", primary visual cortex at the bottom. Likely, the only place where the hierarchical approximation is useful, is when analyzing an isolated, unimodal sensory processing stream. Thus, any serious candidate for an algorithm of cortical function must be implementable on a non-hierarchical network of nodes with symmetric interactions and arbitrary graph topology. JEPA – in its current variants – has been limited to hierarchical networks. We speculate that an interesting way forward would be to develop ways to implement a JEPA-like algorithm on a non-hierarchical network topology, as can be done for predictive processing (Salvatori et al., 2022). Inspired by JEPA and predictive processing, here we outline a set of ideas that that we think would constitute a foundation for a new model of cortical function. This proposal is informed by data, and our emphasis is on the elements that can be experimentally tested.

1. We propose that, similar to a JEPA, cortex is computationally split into two parallel networks embedded in deep, and superficial cortical layers respectively. Such a split is consistent with the dual-counterstream anatomy of cortex (Markov et al., 2013) and dense within layer connectivity (MICrONS Consortium, 2025). In this configuration, representations of superficial network are used to continuously predict activity in the deep cortical network to guide plasticity in deep layers.
2. Superficial cortical layers implement a predictive processing like algorithm with explicit prediction error neurons. However, instead of computing a difference between an external teaching input and a prediction, these error neurons compute a difference between the local representation in layer 5 and a prediction signal. Most prior evidence we have for predictive processing in cortex comes from evidence of prediction error responses in layer 2/3 and is consistent with this (Attinger et al., 2017; Fiser et al., 2016; Garner and Keller, 2022; Jordan and Keller, 2023, 2020; Keller et al., 2012; Leinweber et al., 2017; O'Toole et al., 2023; Solyga and Keller, 2025; Widmer et al., 2022; Zmarz and Keller, 2016). Based on this, we propose that two genetically identified subtypes of layer 2/3 neurons – Rrad and Adamts2 – function as positive and negative prediction error neurons respectively (O'Toole et al., 2023), while a third type of layer 2/3 neuron – Agmat – or a subset of layer 4 could function as internal representation neurons within superficial layers.
3. While the computational architecture of the deep cortical network is experimentally only poorly constrained, we speculate it could also be based on predictive processing. There is an intriguing mirroring of the layer 2/3 response profile in layer 5. Negative prediction error neurons are primarily found in superficial parts of layer 2/3 (O'Toole et al., 2023), likely what is layer 2 in higher mammals, while positive prediction error responses are enriched in deeper parts of layer 2/3. This is similar in layer 5 with strong visuomotor mismatch responses in the more superficial layer 5 IT population (Heindorf and Keller, 2024). It is conceivable that deep and superficial cortical layers might form a predictive processing like architecture, and that superficial layers implement predictive processing on latent representations of the deep layers. In this case superficial layers would constitute a predictive processing module that would be hierarchically above the deep layers, and would allow learning of multiple relational maps, e.g. learning of multiple task contingencies or generalizing across related tasks.
4. The source of external input (Figure 9) to both networks is thalamic projections. Both deep and superficial layers of cortex receive thalamic input (Constantinople and Bruno, 2013; Douglas et al., 1989). Theoretical work inspired by this arrangement also showed that such a network trained in a self-supervised manner would exhibit functional features consistent with

those of cortical responses (Nejad et al., 2025).

The strength of this proposal lies in its integration of three key ideas: JEPA for self-supervised learning (Assran et al., 2023), predictive processing to solve the credit assignment problem (Whittington and Bogacz, 2017) and to enable learning on arbitrary graph topologies (Salvatori et al., 2022), and internal models for motor control (Jordan and Rumelhart, 1992; Kawato, 1999).

Testable predictions

A useful model of a neuronal circuit is likely characterized by three principles:

1. The model is based on physiology and anatomy data.
2. It makes a concrete circuit proposal based on genetically identified cell types.
3. It can be implemented as an algorithm in silico to perform a useful computation.

These principles should enforce sufficient rigor to ensure that models are both informative of how the neuronal circuit might work and remain falsifiable. One of the few theories of cortical function that exhibited most of these characteristics is predictive processing. It was based on previously unexplained observations (Rao and Ballard, 1999), was consistent with the known cortical anatomy of the time (Bastos et al., 2012), there are concrete circuit proposals based on genetically identified cell types (Keller and Mrsic-Flogel, 2018), and it has some limited computational use when implemented (Lotter et al., 2016; Straka et al., 2020). However, given our results here, we conclude that predictive processing as formulated in its most prevalent variety (Keller and Mrsic-Flogel, 2018; Rao and Ballard, 1999) is inadequate to explain cortical computations.

The idea of a JEPA again appears like an opportune starting point to develop the next iteration of a hypothesis for the algorithm of cortex. It shares many of the circuit elements with predictive processing, which have been partially already confirmed to exist in layer 2/3, has computational usefulness, and would be consistent with our results here. There are several experimentally testable predictions we can make based on the idea that cortex implements a JEPA-like computation as detailed in the previous section:

1. Assuming superficial cortical layers implement one of two JEPA networks that only communicate with the network in deep cortical layers, but do not project directly out of cortex, then silencing layer 2/3 should interfere with learning in layer 5 but leave activity in layer 5, and behavior more generally, otherwise largely unaffected. This is in stark contrast to predictions the canonical microcircuit model would make.
2. A JEPA would also predict a difference in learning rates between the encoder networks and the predictor network (Assran et al., 2023; Grill et al., 2020; Mohammadi et al., 2025). Based on this, we should find higher levels of plasticity in the layer 2/3 prediction error network, than in layer 4 or the corresponding representation network in deep cortical layers.
3. There should be a neuron type in layer 2/3 or layer 4 that acts as an internal representation neuron for the prediction error neurons in layer 2/3. This could either be Agmat positive layer 2/3 neurons (O'Toole et al., 2023), or a subset of layer 4 neurons. These neurons should exhibit opposing functional influence patterns on prediction error neurons as predicted by the predictive processing circuit model (Keller and Mrsic-Flogel, 2018).

We believe that testing predictions outlined above is a promising way forward to further constrain the search for a candidate algorithm of cortical function and formulate a more complete proposal.

Methods

Animals

All animal procedures were approved by and carried out in accordance with guidelines of the Veterinary Department of the Canton Basel-Stadt, Switzerland. The mice used in this study were: 23 Tlx3-Cre (Gerfen et al., 2013), 6 Fezf2-CreERT2 (Matho et al., 2021), 6 Ntsr1-Cre (Gong et al., 2007), 4 Scnn1a-Cre (Madisen et al., 2010) and 8 Tlx3-Cre x Ai148 (Daigle et al., 2018).

Transgenic mice were bred in-house and heterozygous mice of either sex were used. 8 females of C57BL/6J (Charles River Laboratories) were used for control experiments. All mice were aged between 7 and 22 weeks. Mice were co-housed in groups of 2 to 5 in a reversed 12 hour light-dark cycle and had ad libitum access to water and regular mouse food. To induce expression of Cre in Fezf2-CreERT2 mice, their regular food was substituted for tamoxifen containing food (Envigo, 400 mg tamoxifen per kg of food) for the duration of 3 weeks. All imaging experiments took place in the dark phase of the light cycle.

Surgery and viral vector injections

For all surgical procedures, mice were anesthetized using a mix of fentanyl (0.05 mg/kg), medetomidine (0.5 mg/kg) and midazolam (5 mg/kg). Analgesics were applied peri- and postoperatively. Lidocaine was injected locally on the scalp (10 mg/kg s.c.) prior to surgery. To provide postoperative analgesia we used either a combination of repeated injections of metacam (5 mg/kg, s.c.) and buprenorphine (0.1 mg/kg s.c.) or extended-release buprenorphine (Ethiqa XR, 3.25 mg/kg s.c.). Mice were returned to their home cage after the anesthesia was antagonized by a mix of flumazenil (0.5 mg/kg) and atipamezole (2.5 mg/kg) and were allowed to recover for at least one week prior to first head-fixation.

For two-photon imaging and concurrent optogenetic stimulation experiments, mice underwent a cranial window implantation surgery. First, a circular craniotomy 4 mm in diameter was made over right V1 centered on 2.5 mm lateral and 0.5 mm anterior of lambda, which was followed by a durectomy. Then an AAV vector mix was injected (approximately 300 nl per injection) across 4-6 locations spanning right V1. For mapping the functional influence from cell types of deep layers on layer 2/3, the AAV vector mix consisted of AAV2/1-EF1α-GCaMP6f (5×10^{12} - 4×10^{14} GC/ml) and AAV2/1-hSyn-DIO-ChrimsonR-tdTomato (10^{13} – 3×10^{14} GC/ml). For no-ChrimsonR control mice, the ChrimsonR vector was omitted. For mapping the functional influence from cell types of layer 2/3 onto layer 5, Tlx3-Cre x Ai148 mice were injected with either AAV2/1-AP.Baz1a.1-CreOFF-ChrimsonR-mRuby-ST (4.6×10^{13} GC/ml) or AAV2/1-AP.Adamts2.1-CreOFF-ChrimsonR-mRuby-ST (2.5×10^{13} GC/ml). For assessing the responses of Tlx3 neurons to optogenetic activation of the same population, mice were injected with AAV2/9-DIO-jGCaMP8s-P2A-ChrimsonR-ST (1.1×10^{15} GC/ml). A 4 mm diameter glass window was then placed over the craniotomy and sealed with cyanoacrylate. Lastly, a custom titanium headplate was attached to the skull using dental cement (Paladur, Heraeus).

Virtual reality and visuomotor paradigms

During all recordings, mice were head-fixed in a virtual reality system as described previously (Leinweber et al., 2014). Briefly, mice were free to run on an air-supported polystyrene ball of 20 cm in diameter, and the virtual environment was projected onto a toroidal screen surrounding the mouse. From the point of view of the mouse, the screen covered a visual field of approximately 240 degrees horizontally and 100 degrees vertically. The projector output was gated by a 24 kHz gating signal to only turn the illumination on at the turnaround points of the two-photon resonance scanner. This was done to minimize light leak from the virtual reality system. In the closed loop and open loop conditions, the virtual environment presented on the screen was a virtual tunnel with walls textured with vertical sinusoidal gratings. Movement in the virtual tunnel was restricted to one dimension (forward and backwards). In the closed loop condition, visual flow speed experienced by the mouse was coupled to its locomotion speed, with the exception of brief (1 s) mismatch events presented every 12 ± 4.5 s (mean ± SD). During these events visual flow speed was clamped to zero for 1 s. In the open loop condition, the visual flow presented was a playback of the visual flow the mouse had self-generated in the preceding closed loop session. For the grating sessions, we presented the mice with full-field sinusoidal drifting gratings at 0 degrees, 45 degrees, 90 degrees and 135 degrees orientations, moving in either direction. Each stimulus in a sequence was selected randomly and presented for a duration of 6 ± 1.35 s (mean ± SD) with a randomized inter-stimulus interval of 4.5 ± 1 s (mean ± SD) during which a gray screen was displayed. For closed loop and open loop optogenetic activation of layer 5

experiments (Figure 7), mice were presented with a static gray screen. Prior to the imaging experiments, mice were habituated to head-fixation and locomotion on the ball during daily 1-hour training sessions for at least 3 days, until they displayed regular locomotion.

Two-photon calcium imaging

Two-photon calcium imaging was performed using a modified Thorlabs Bergamo II two-photon microscope. Excitation illumination was provided by a titanium-sapphire laser (Mai Tai, Spectra-Physics) tuned to 930 nm. The emission light was band-pass filtered using a 525/50 filter (Semrock) and detected by a photomultiplier tube (H7422, Hamamatsu). The detected signal was amplified (DHPCA-100, Femto), digitized (NI5772, National Instruments) and band-pass filtered at 80 MHz using a digital Fourier-transform filter implemented on an FPGA (NI5772, National Instruments) using custom-written software. A 12 kHz scanning system (Cambridge Technology) was used to acquire images at 750 x 400 pixels resolution at 60 Hz frame rate. The field of view under the microscope was approximately 300 x 300 μm. A piezo electric linear actuator (P726 PIFOC, Physik Instrumente) was used to move the objective (Nikon 16x, 0.8 NA) to image 4 separate planes sequentially, which resulted in an effective frame rate of 15 Hz. Custom software was used to acquire imaging data.

Concurrent two-photon imaging and optogenetic stimulation

Illumination source for the optogenetic stimulation of ChrimsonR-expressing neurons was a 637 nm laser (OBIS, Coherent). The laser beam was merged with the main imaging path using a dichroic mirror (ZT775sp-2p, Chroma) and sent through the two-photon microscope objective. Another dichroic mirror (F38-555SG, Semrock) was used to split the GFP emission from both illumination light sources. To reduce the light leak artifacts during optogenetic stimulation, the laser output was synchronized to the turnaround times of the resonant scanner. An essential addition to artifact free stimulation was the digital band-pass filter used to process the PMT signals (see above). To optogenetically activate cell types of layers 4, 5 and 6 (Figures 3-5 and 8), we used contentious laser pulses of 1 s duration, presented randomly in closed loop and open loop sessions every 12 ± 3.5 s. The laser power used for optogenetic stimulation in these experiments was fixed at 11 mW. For optogenetic activation of layer 2/3 cell types (Figure 6) stimulation occurred on average once every 11 s, and the stimulation power varied between 1 mW and 11 mW. The responses were then averaged over all stimulation powers.

Closed loop and open loop optogenetic activation of layer 5

For the closed loop optogenetic activation experiment (Figure 7) the stimulation laser power was coupled to the locomotion speed of the mouse. The stimulation power scaled linearly with the speed of the mouse and ranged from the minimum of 0 mW (when mice were stationary), to a maximum of 6 mW (set to the maximum locomotion speed of the preceding sessions). The opto-motor mismatch events were triggered by clamping the stimulation power to 0 mW for 1 s at random times every 12 ± 3.5 s (mean ± SD). The open loop condition consisted of a replay of the stimulation power profile the mouse had self-generated in the preceding closed loop session including optomotor mismatch events.

Extraction of neuronal activity

Functional two-photon calcium imaging data were processed as previously described (Keller et al., 2012). Briefly, raw images were full-frame registered to correct for lateral brain motion. Neurons were selected manually based on mean and maximum fluorescence images. Fluorescence traces of each neuron were calculated as the mean fluorescence over all pixels in the given region of interest and corrected for slow drift in fluorescence using an 8^{th}-percentile filter and 1000 frame (67 s) window. For $\Delta F/F_0$ calculation, F_0 was defined as the median fluorescence over the entire trace.

Data analysis

To quantify the average response traces, we used a hierarchical bootstrap approach (Saravanan et al., 2020) to estimate the mean activity value at each time bin. Briefly, for each trace the data were first resampled with replacement at the level of recording sites. From the sampled set of recording sites, the data were then resampled with replacement at the level of neurons. For each of 10 000 such bootstrap samples we calculated the mean, to generate a distribution of bootstrap mean values. We used the mean of this distribution as our estimate of the true mean, and the standard deviation of this distribution as bootstrap standard error. All average response traces were then baseline (mean $\Delta F/F_0$ in 0.6 s window prior to the stimulus onset) subtracted.

For the analysis of visuomotor and optomotor mismatch responses, we included only mismatch events that occurred during locomotion periods – when the average locomotion velocity in the window -1 to +1 s around the mismatch onset was above 2.7 cm/s. Visuomotor mismatch events that overlapped with optogenetic stimulation in the interval of -1 to +1 s from mismatch onset were excluded from the analysis. For the analysis of grating onset responses, we included only grating onsets that occurred during stationary periods (when the average locomotion velocity in the -1 to +1 s window around grating onset was below 1.8 cm/s). The only exception was the analysis of responses to a specific grating orientation (Figures S5A and S5B), for which all trials were included due to a low number of repetitions per orientation. The same windows and thresholds were used to classify locomotion and stationary periods for the analysis of optogenetic stimulation responses (Figure S8). Locomotion onset was defined as a time point at which locomotion speed crossed a threshold of 1.8 cm/s, provided that the mean speed in the preceding 1 s was below the threshold and the mean speed in the following 1 s remained above this threshold. For analysis of locomotion onsets during open loop or grating sessions, locomotion onsets that occurred within a window -0.33 to + 2 s relative to a visual stimulus onset were excluded.

Neuronal responses to visuomotor mismatch, grating onset, optomotor mismatch and optomotor open loop stimulation offset were calculated as the mean activity of each neuron in a 0.33 to 1.33 s response window after stimulation onset minus the mean activity of each neuron in a 0.6 s baseline window preceding the stimulation onset. For quantification of neuronal responses to optogenetic activation, a response window of 0.2 to 1.33 s was used to account for faster response dynamics. To identify positive and negative prediction error neurons, neurons were classified based on their responses to grating onset and visuomotor mismatch, respectively. For each stimulus, the top 15% (Figures 2-5, 8, S1, S3-S7, S9), 10% or 20% (Figures S3, S6, S7, S9) of responsive neurons were selected ensuring no overlap. For all analyses in which we select and plot on the same variable (Figures 2-6, 8, S5), we selected neurons based on half the trials and used the other half of the trials for plotting. This was done to avoid one type of regression to the mean artifact that results from selecting the most responsive neurons to a given stimulus and then plotting the responses of these neurons based on the same data. To prevent graphical aliasing, all heatmaps were smoothed across 3 to 30 neurons, with the smoothing window scaled to the size of the plotted population to ensure correct display at minimum resolution of 210 pixels per heatmap. To determine the correlation between neuronal activity and optical flow speed or locomotion speed during the grating sessions, we calculated the correlation coefficient between the individual neuronal activity traces and binarized grating trace or activity traces and locomotion speed respectively (Figures 2-5 and 8). To quantify the neuronal responses to optomotor mismatch as a function of visuomotor mismatch response (Figure 7), we binned layer 2/3 neurons in 5 groups based on their visuomotor mismatch response. The bins comprised 0th-10th, 10th-25th, 25th-85th, 85th-95th and 95th-100th percentiles of responses distribution. The top two bins were chosen to reflect the top 15% of neurons used for analysis elsewhere, the rest were chosen pseudo randomly. Exact choice of bins did not influence our conclusions.

Statistical tests

All statistical information for the tests performed in this manuscript is provided in Table S1. For statistical testing we used a hierarchical bootstrapping approach to account for the nested structure of the data. Each recording site contains a different number of neurons, so neurons (lower level of the hierarchical data structure) are grouped within the recording sites (higher level of the hierarchical data structure). Data are resampled hierarchically to generate bootstrap estimates (Saravanan et al., 2020). For the quantification of average response traces, the data were first resampled with replacement at the level of recording sites and then resampled with replacement at the level of neurons. For each of 10 000 such bootstrap samples we calculated the mean to generate a bootstrap distribution of mean values. The p-value was then calculated as a fraction of the bootstrap sample means which violated the tested hypothesis. To report the difference between two average calcium traces (Figures 2-8, S1, S3-S9), we performed hierarchical bootstrapping for every time bin of the calcium trace and marked bins where responses were significantly different ($p < 0.05$). We removed isolated significant bins, such that only intervals of at least two consecutive significant bins were marked.

Supplementary material

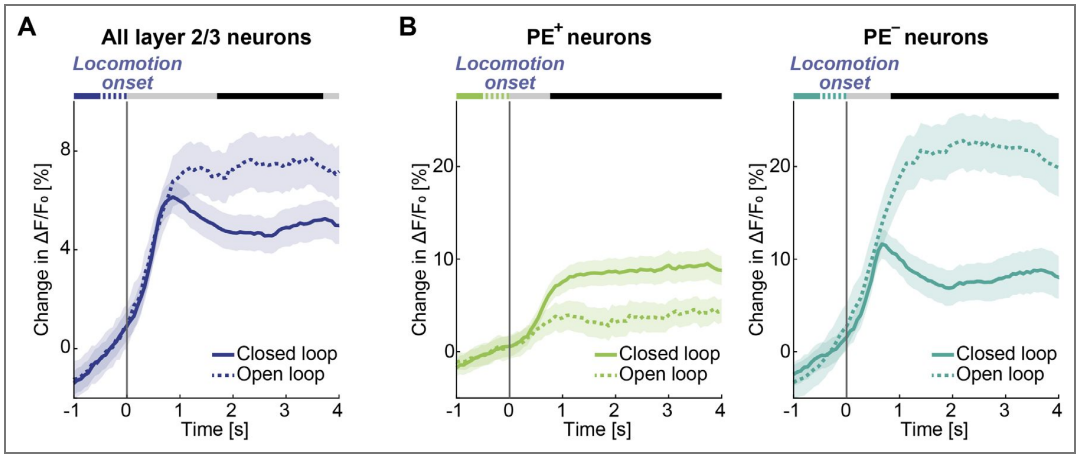


Figure S1. Locomotion onset responses of layer 2/3 neurons differ between closed loop and open loop conditions. (A) Mean layer 2/3 population response to locomotion onset in closed loop (solid line) and open loop (dashed line) conditions. Here and in the subsequent panels, response curves represent the hierarchical bootstrap estimates of the mean trace. Shading around the line is the bootstrap error, defined as one standard deviation of the bootstrap distribution at each time bin. Response curves are compared for each time bin: in the horizontal bars above the plot, black marks time bins where $p<0.05$ and gray mark time bins where $p>0.05$. **(B)** Mean response of positive (left, light green) and negative (right, dark green) prediction error neurons to locomotion onset in closed loop (solid lines) and open loop (dashed lines) conditions.

Figure S2. Responses of layer 2/3 neurons to control light stimulation do not correlate with optogenetic stimulation responses and resemble responses in no-ChrimsonR control mice.

(A) Scatterplot of the responses of layer 2/3 neurons to optogenetic stimulation of Tlx3 layer 5 neurons against their responses to control light stimulation. Each dot represents a neuron. **(B)** Comparison of average responses of layer 2/3 neurons to control light stimulation (yellow) and stimulation light in no-ChrimsonR control mice (light gray). Boxes indicate the interquartile range, the central line marks the median, and whiskers the 10th and 90th percentiles. Orange circles mark the mean response.

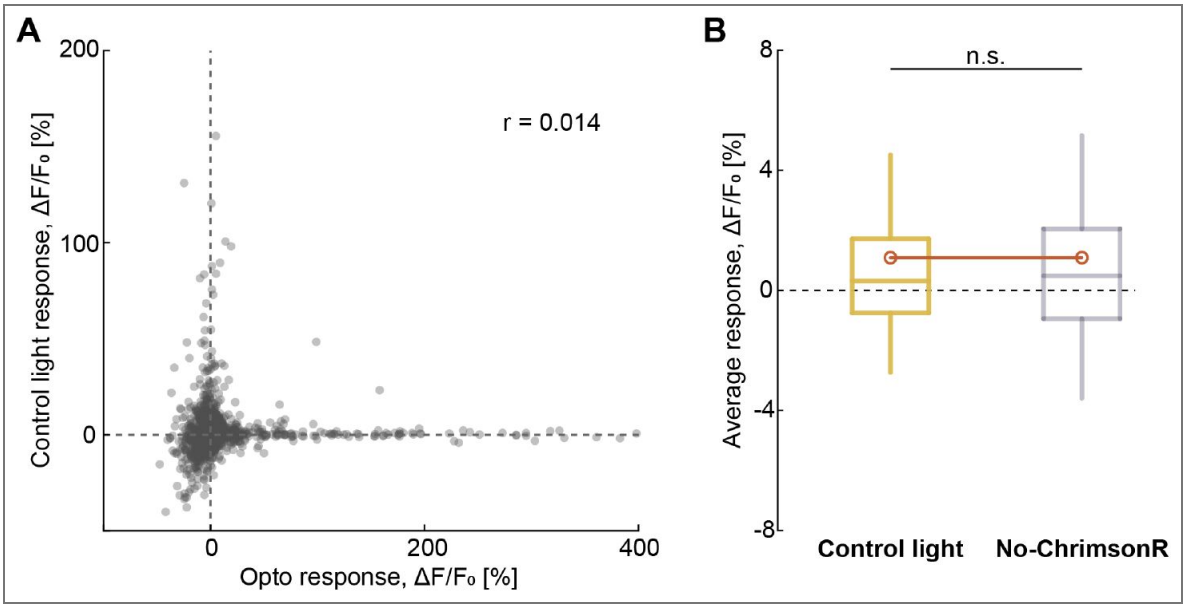
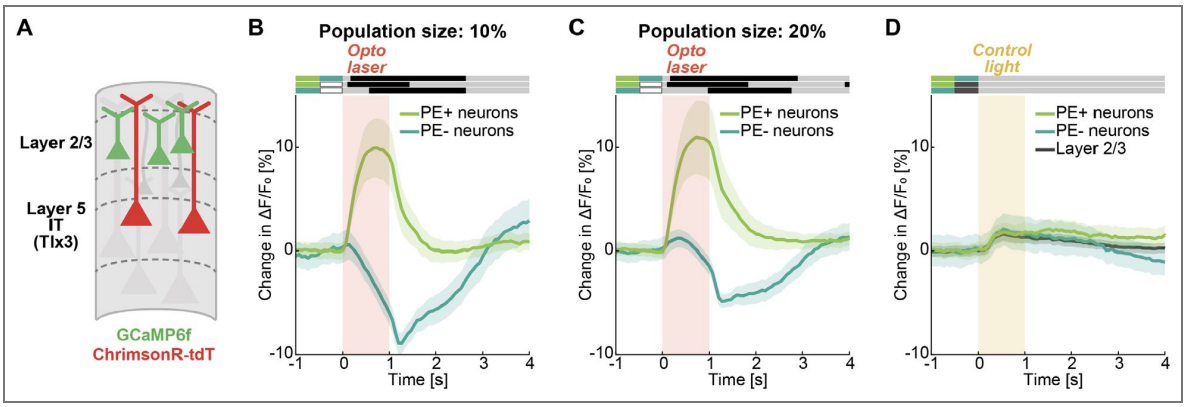


Figure S3. Functionally specific modulation by Tlx3 layer 5 neurons is robust to functional group size and is absent for control light stimulation.

(A) We optogenetically activated Tlx3 layer 5 neurons while recording calcium activity in layer 2/3 neurons. **(B)** As in Figure 3D, but using the 10% (instead of 15%) most strongly responding neurons to grating onset and visuomotor mismatch. **(C)** As in Figure 3D, but using the 20% (instead of 15%) most strongly responding neurons to grating onset and visuomotor mismatch. **(D)** Mean responses of positive prediction error neurons (light green), negative prediction error neurons (dark green), and the layer 2/3 population (black) to control light stimulation. Yellow shading marks the control stimulation interval.



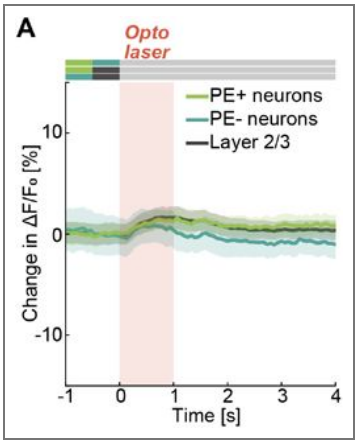


Figure S4. No-ChrimsonR control stimulation does not induce functionally specific modulation of layer 2/3 neurons.

(**A**) Mean responses of positive prediction error neurons (light green), negative prediction error neurons (dark green), and the layer 2/3 population (black) to stimulation light in no-ChrimsonR control mice. Red shading marks the stimulation interval.

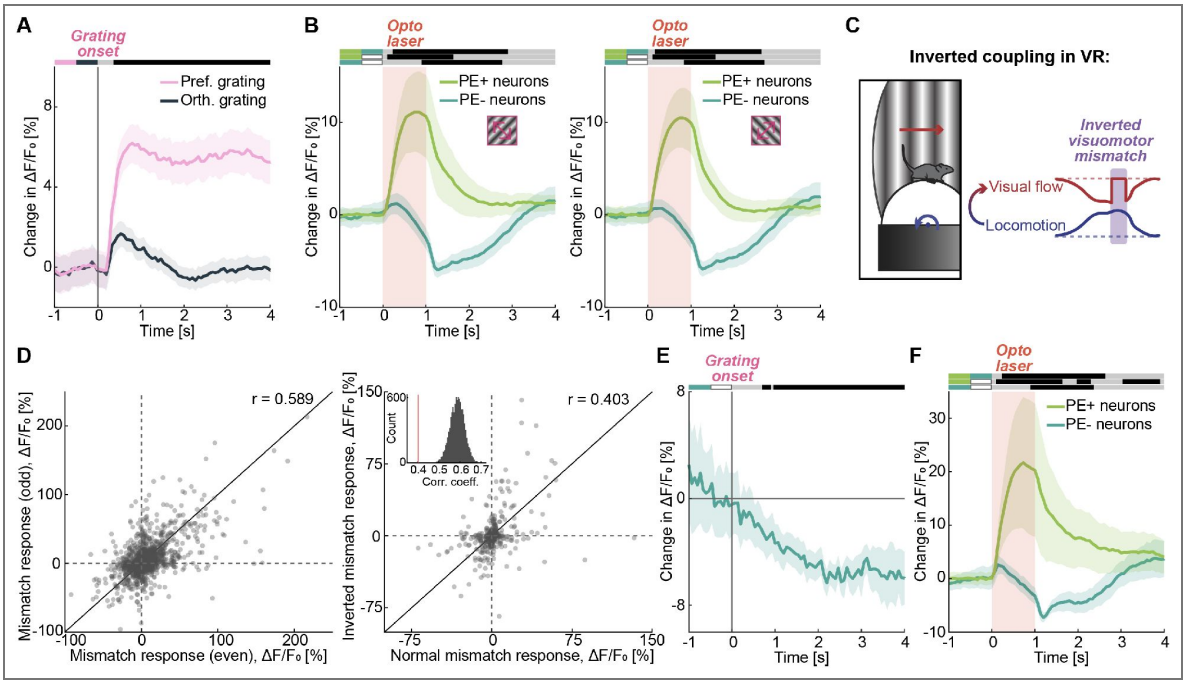


Figure S5. Stimulation of layer 5 likely separates positive and negative error neurons irrespective of feature-selectivity of those neurons.

(A) Mean layer 2/3 population response to preferred grating (pink) and orthogonal grating (black) onsets. **(B)** Mean responses of positive prediction error neurons (light green) and negative prediction error neurons (dark green) to optogenetic stimulation of Tlx3 layer 5 neurons. Positive prediction error neurons were selected using a single grating orientation, either 135° (left) or 45° (right). Red shading marks the stimulation interval. **(C)** Schematic of the inverted visuomotor coupling in the virtual reality system. Locomotion of the mouse on the treadmill is coupled to forward moving visual flow. We define inverted visuomotor mismatch stimulus as a halt in the inverted visual flow during locomotion of the mouse. **(D)** Inverted visuomotor mismatch drives a population of neurons different from those activated by normal visuomotor mismatch. Left: quantification of measurement noise in normal visuomotor mismatch responses. Scatter plot of visuomotor mismatch response on even trials versus the same response on odd trials. Each dot is a neuron. The black line marks the diagonal. Right: Responses of layer 2/3 neurons to normal visuomotor mismatch versus their responses to inverted visuomotor mismatch. The correlation between the two is significantly lower than would be expected from measurement noise. Inset: Bootstrap distribution of odd versus even trial correlations of normal visuomotor mismatch responses. Vertical red line marks the observed correlation between normal and inverted visuomotor mismatch responses. **(E)** Mean response of inverted visuomotor mismatch neurons to grating onset. Like normal visuomotor mismatch neurons (Figure 2F), inverted visuomotor mismatch neurons decrease their activity on visual stimulation. **(F)** Mean responses of positive prediction error neurons (light green) and negative prediction error neurons (dark green) to optogenetic stimulation of Tlx3 layer 5 neurons. Negative prediction error neurons are selected using inverted visuomotor mismatch stimulus. Red shading marks the stimulation interval.

Figure S6. Functionally specific modulation by Fezf2 layer 5 neurons is robust to functional group size and is absent for control light stimulation.

(A) We optogenetically activated Fezf2 layer 5 neurons while recording calcium activity in layer 2/3 neurons. **(B)** As in Figure 4D, but using the 10% (instead of 15%) most strongly responding neurons to grating onset and visuomotor mismatch. **(C)** As in Figure 4D, but using the 20% (instead of 15%) most strongly responding neurons to grating onset and visuomotor mismatch. **(D)** Mean responses of positive prediction error neurons (light green), negative prediction error neurons (dark green), and the layer 2/3 population (black) to control light stimulation. Yellow shading marks the control stimulation interval.

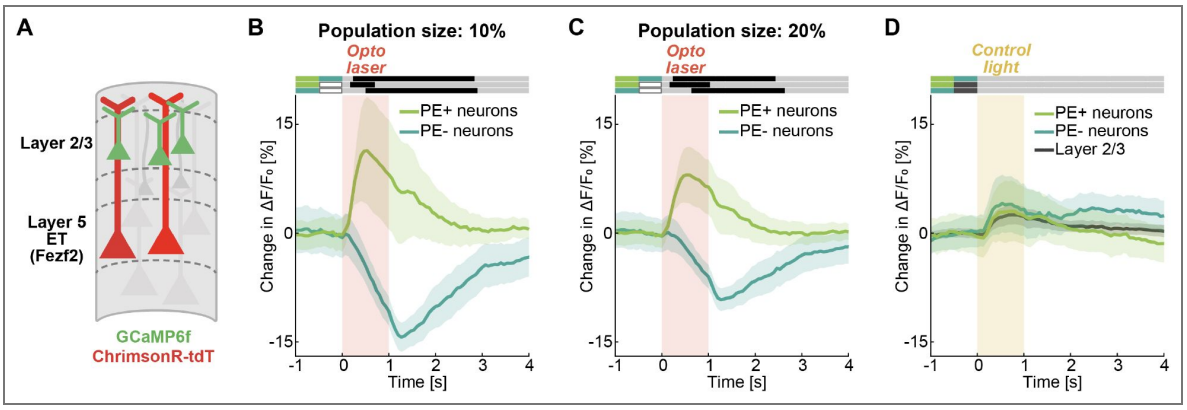
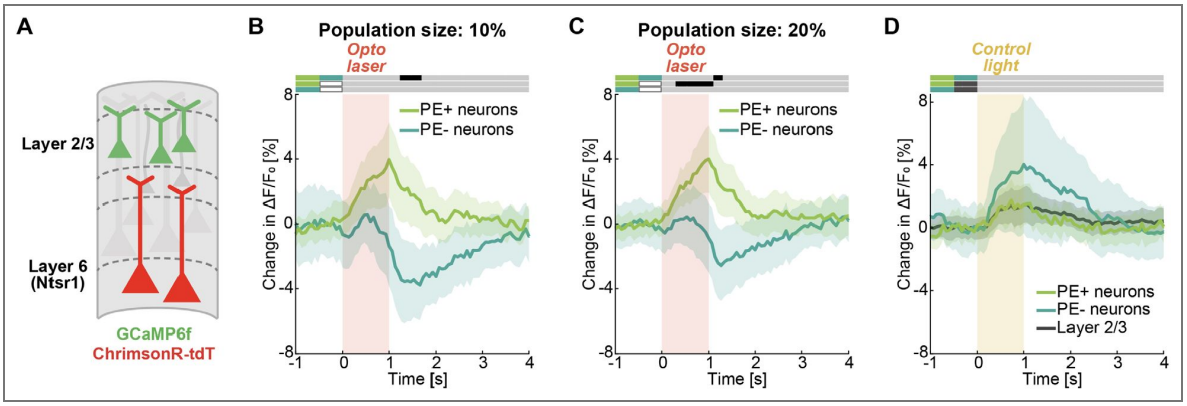


Figure S7. Functionally specific modulation by Ntsr1 layer 6 neurons is robust to functional group size and is absent for control light stimulation.

(A) We optogenetically activated Ntsr1 layer 6 neurons while recording calcium activity in layer 2/3 neurons. **(B)** As in Figure 5D, but using the 10% (instead of 15%) most strongly responding neurons to grating onset and visuomotor mismatch. **(C)** As in Figure 5D, but using the 20% (instead of 15%) most strongly responding neurons to grating onset and visuomotor mismatch. **(D)** Mean responses of positive prediction error neurons (light green), negative prediction error neurons (dark green), and the layer 2/3 population (black) to control light stimulation. Yellow shading marks the control stimulation interval.



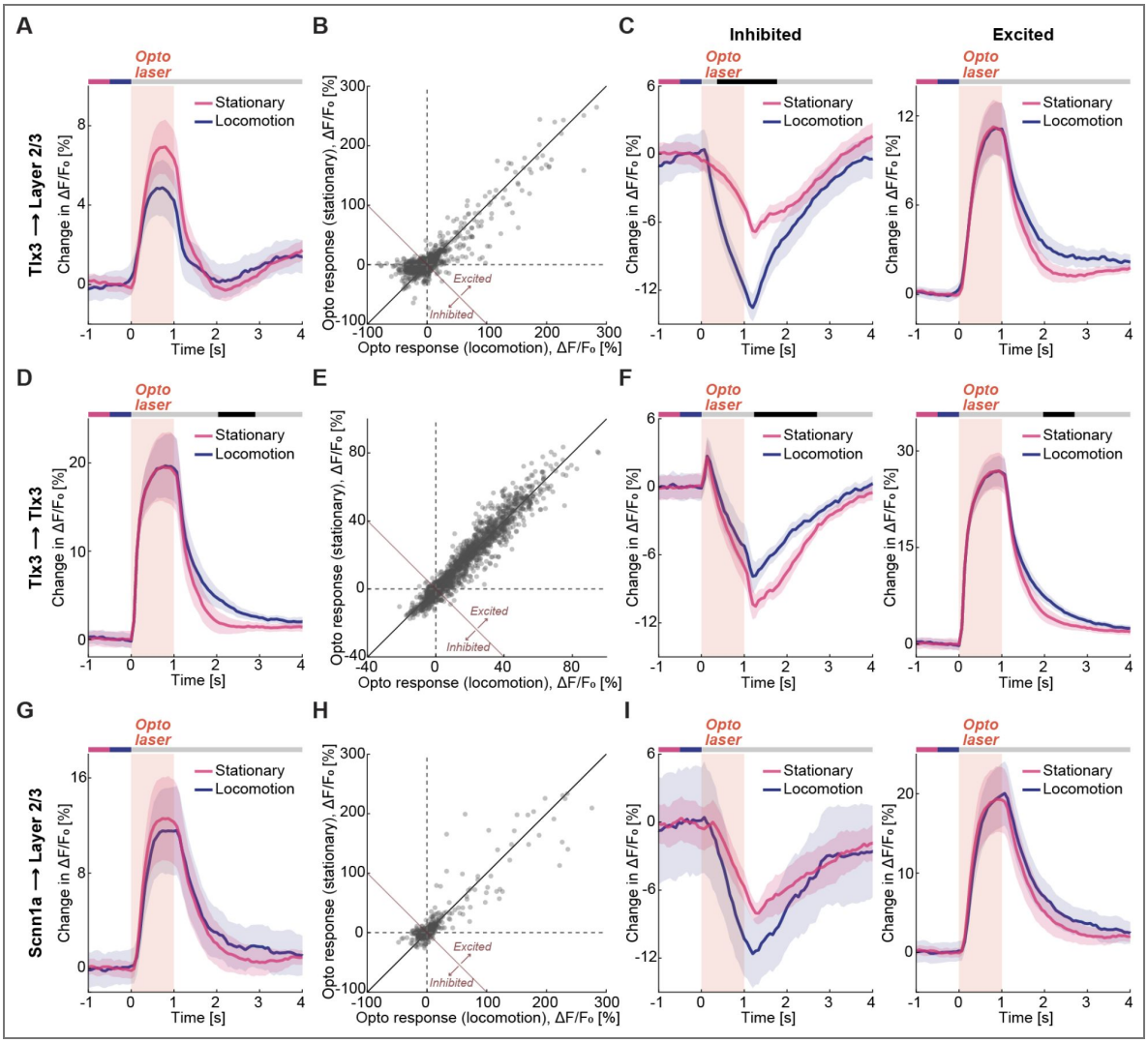


Figure S8. Locomotion unmasks inhibitory influence of layer 5 IT neurons, but not layer 4 neurons, on layer 2/3 neurons.

(A) Mean layer 2/3 population response to optogenetic stimulation of Tlx3 layer 5 neurons split by whether the mouse is stationary (pink) or locomoting (blue). **(B)** Scatterplot of layer 2/3 neuronal responses to stimulation of Tlx3 layer 5 neurons while mice were locomoting versus their responses while mice were stationary. Each dot is a neuron. For the analysis shown in **C**, we define neurons below and to the left of the negative diagonal as inhibited, and those to the right and above as excited. **(C)** Left: Mean response of inhibited layer 2/3 neurons (as defined in **B**). Right: Mean response of excited layer 2/3 neurons. (**D-F**) As in **A-C**, but for responses of Tlx3 layer 5 neurons to optogenetic stimulation of the same population. (**G-I**) As in **A-C**, but for responses of layer 2/3 neurons to optogenetic stimulation of Scnn1a layer 4 neurons.

Figure S9. Neither optogenetic stimulation of Scnn1a layer 4 neurons nor control light stimulation results in functionally specific modulation in layer 2/3.

(A) We optogenetically activated Scnn1a layer 4 neurons while recording calcium activity in layer 2/3 neurons. **(B)** As in Figure 8D, but using the 10% (instead of 15%) most strongly responding neurons to grating onset and visuomotor mismatch. **(C)** As in Figure 8D, but using the 20% (instead of 15%) most strongly responding neurons to grating onset and visuomotor mismatch. **(D)** Mean responses of positive prediction error neurons (light green), negative prediction error neurons (dark green), and the layer 2/3 population (black) to control light stimulation. Yellow shading marks the control stimulation interval.

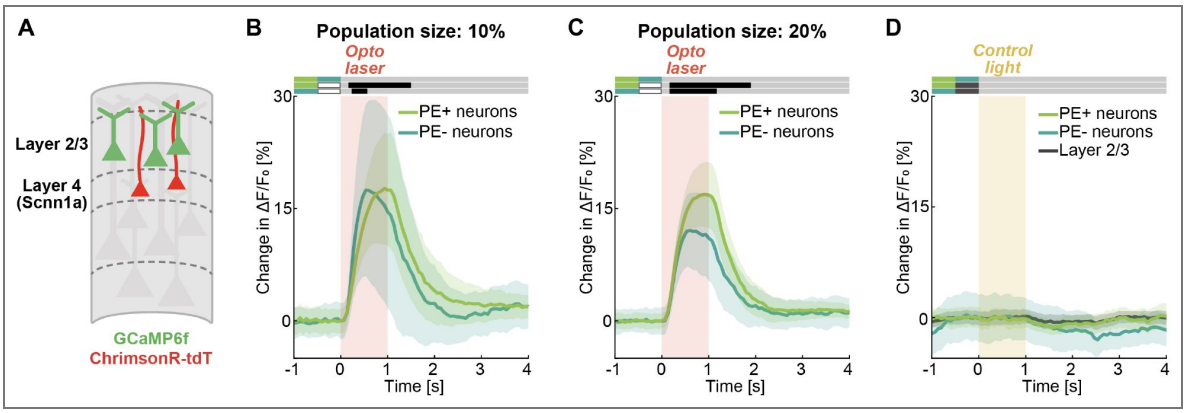
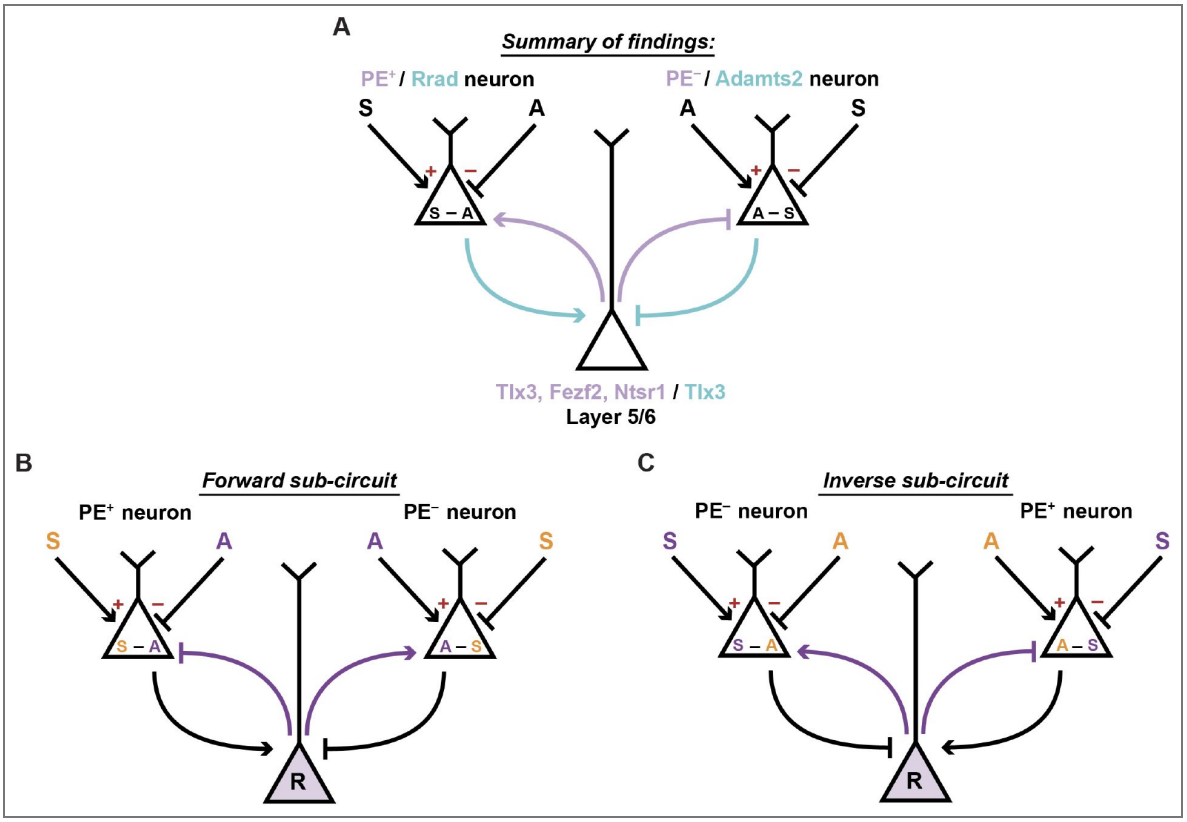


Figure S10. Sub-circuits for reciprocal predictive processing in a single cortical area.

(A) Schematic of the functional influence patterns between cell types of deep and superficial cortical layers. Purple color marks functional connectivity from genetically identified cell types of deep layers on functionally identified prediction error types of layer 2/3. Blue color marks functional connectivity from molecularly identified cell types of layer 2/3 onto Tlx3 neurons of layer 5. **(B)** Schematic of the first (forward) sub-circuit. The action (A) input functions as a prediction and the sensory input (S) as a teaching signal. The internal representation neuron represents an estimate of the sensory input. **(C)** Schematic of the second (inverse) sub-circuit. The action (A) input functions as a teaching signal, and the sensory input (S) as a prediction. The internal representation neuron represents an estimate of the action.



REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
AAV2/1-EF1α-GCaMP6f (10^{12}-10^{14} GC/ml)	FMI vector core	vector.fmi.ch
AAV2/1-hSyn-DIO-ChrimsonR-tdTomato (10^{13}-10^{14} GC/ml)	FMI vector core	vector.fmi.ch
AAV2/1-AP.Baz1a.1-CreOFF-ChrimsonR-mRuby2-ST (10^{13} GC/ml)	FMI vector core	vector.fmi.ch
AAV2/1-AP.Adamts2.1-CreOFF-ChrimsonR-mRuby2-ST (10^{13} GC/ml)	FMI vector core	vector.fmi.ch
AAV2/9-DIO-jGCaMP8s-P2A-ChrimsonR-ST (10^{15} GC/ml)	FMI vector core	vector.fmi.ch
Chemicals, peptides, and recombinant proteins		
Fentanyl citrate	Actavis	CAS 990-73-8
Midazolam (Dormicum)	Roche	CAS 59467-96-8
Medetomidine (Domitor)	Orion Pharma	CAS 86347-14-0
Ropivacaine	Presenius Kabi	CAS 132112-35-7
Lidocaine	Bichsel	CAS 137-58-6
Buprenorphine	Reckitt Benckiser Healthcare	CAS 52485-79-7
Ophthalmic gel (Humigel)	Virbac	N/A
Flumazenil (Anexate)	Roche	CAS 78755-81-4
Atipamezole (Antisedan)	Orion Pharma	CAS 104054-27-5
N-Butyl-2-cyanoacrylate (Vetbond)	3M	CAS 6606-65-1
Dental cement (Paladur)	Heraeus Kulzer	CAS 9066-86-8
Metacam	Boehringer Ingelheim	CAS 71125-39-8
Buprenorphine hydrochloride (Ethiqa XR)	Fidelis Pharmaceuticals	CAS 53152-21-9
Tamoxifen food	Envigo	CAS 10540-29-1 (TD.55125)
Deposited data		
All data and code used to generate manuscript figures	This paper	zenodo.org
Experimental models: Organisms/strains		
Mus musculus: C57BL/6	Charles River	N/A
Mus musculus: Tg(Tlx3-cre)PL56Gsat/Mmucd	MMRRC	RRID:MMRRC_041158-UCD
Mus musculus: Fezf2$^{tm1.1(cre/ERT2)Zjh}$/J	Jackson Laboratories	RRID:IMSR_JAX:036296
Mus musculus: Tg(Ntsr1-cre)GN220Gsat/Mmucd	MMRRC	RRID:MMRRC_017266-UCD
Mus musculus: Tg(Scnn1a-cre)3Aibs/J	Jackson Laboratories	RRID:IMSR_JAX:009613
Mus musculus: Igs7$^{tm148.1(tetO-GCaMP6f,CAG-tTA2)Hze}$/J	Jackson Laboratories	RRID:IMSR_JAX:030328
Software and algorithms		
MATLAB (2023b)	The MathWorks	RRID:SCR_001622
LabVIEW	National Instruments	RRID:SCR_014325
Python	python.org	RRID:SCR_008394
Panda3D	panda3d.org	N/A

Key resources table.

Reference	Description	Test	N_1 (sites/mice)	N_2 (sites/mice)	Unit	P-value
Figure 2D	Mean layer 2/3 response to grating	-	7030 (70/36)	-	Neurons	-
Figure 2E	Mean layer 2/3 response to MM	-	7030 (70/36)	-	Neurons	-
Figure 2F	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	1056 (69/36)	1056 (69/36)	Neurons	see Figure
Figure 2G	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	1056 (69/36)	1056 (69/36)	Neurons	see Figure
Figure 2H	Correlation between activity and grating and activity and locomotion speed	-	7030 (70/36)	-	Neurons	-
Figure 3B	Mean response ChrimsonR (N_1) vs Control stim (N_2)	Hierarchical bootstrap	4262 (42/19)	4262 (42/19)	Neurons	see Figure
	Mean response ChrimsonR (N_1) vs no-ChrimsonR (N_2)	Hierarchical bootstrap	4262 (42/19)	1394 (14/8)	Neurons	see Figure
	Mean response Control stim (N_1) vs no-ChrimsonR (N_2)	Hierarchical bootstrap	4262 (42/19)	1394 (14/8)	Neurons	see Figure
Figure 3D	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	539 (35/17)	539 (35/17)	Neurons	see Figure
	Mean response PE+ vs 0	Hierarchical bootstrap	539 (35/17)	-	Neurons	see Figure
	Mean response PE- vs 0	Hierarchical bootstrap	539 (35/17)	-	Neurons	see Figure
Figure 3E	Mean response PE+ (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	539 (35/17)	3591 (36/17)	Neurons	0.0895
	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	539 (35/17)	539 (35/17)	Neurons	$<10^{-4}$
	Mean response layer2/3 (N_1) vs PE- (N_2)	Hierarchical bootstrap	3591 (36/17)	539 (35/17)	Neurons	$<10^{-4}$
	Mean response PE+ vs 0	Hierarchical bootstrap	539 (35/17)	-	Neurons	$<10^{-4}$
	Mean response layer2/3 vs 0	Hierarchical bootstrap	3591 (36/17)	-	Neurons	$<10^{-4}$
	Mean response PE- vs 0	Hierarchical bootstrap	539 (35/17)	-	Neurons	0.0288
Figure 3F	Correlation between activity and grating and activity and locomotion speed	-	3591 (36/17)	-	Neurons	-
Figure 4B	Mean response ChrimsonR (N_1) vs Control stim (N_2)	Hierarchical bootstrap	1494 (13/6)	1494 (13/6)	Neurons	see Figure
	Mean response ChrimsonR (N_1) vs no-ChrimsonR (N_2)	Hierarchical bootstrap	1494 (13/6)	1394 (14/8)	Neurons	see Figure
	Mean response Control stim (N_1) vs no-ChrimsonR (N_2)	Hierarchical bootstrap	1494 (13/6)	1394 (14/8)	Neurons	see Figure
Figure 4D	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	225 (13/6)	225 (13/6)	Neurons	see Figure
	Mean response PE+ vs 0	Hierarchical bootstrap	225 (13/6)	-	Neurons	see Figure
	Mean response PE- vs 0	Hierarchical bootstrap	225 (13/6)	-	Neurons	see Figure
Figure 4E	Mean response PE+ (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	225 (13/6)	1494 (13/6)	Neurons	0.3343
	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	225 (13/6)	225 (13/6)	Neurons	0.0006
	Mean response layer2/3 (N_1) vs PE- (N_2)	Hierarchical bootstrap	1494 (13/6)	225 (13/6)	Neurons	$<10^{-4}$
	Mean response PE+ vs 0	Hierarchical bootstrap	225 (13/6)	-	Neurons	0.2454
	Mean response layer2/3 vs 0	Hierarchical bootstrap	1494 (13/6)	-	Neurons	0.0477
	Mean response PE- vs 0	Hierarchical bootstrap	225 (13/6)	-	Neurons	$<10^{-4}$
Figure 4F	Correlation between activity and grating and activity and locomotion speed	-	1494 (13/6)	-	Neurons	-
Figure 5B	Mean response ChrimsonR (N_1) vs Control stim (N_2)	Hierarchical bootstrap	918 (12/6)	918 (12/6)	Neurons	see Figure
	Mean response ChrimsonR (N_1) vs no-ChrimsonR (N_2)	Hierarchical bootstrap	918 (12/6)	1394 (14/8)	Neurons	see Figure
	Mean response Control stim (N_1) vs no-ChrimsonR (N_2)	Hierarchical bootstrap	918 (12/6)	1394 (14/8)	Neurons	see Figure
Figure 5D	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	131 (11/6)	131 (11/6)	Neurons	see Figure
	Mean response PE+ vs 0	Hierarchical bootstrap	131 (11/6)	-	Neurons	see Figure
	Mean response PE- vs 0	Hierarchical bootstrap	131 (11/6)	-	Neurons	see Figure
Figure 5E	Mean response PE+ (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	131 (11/6)	870 (11/6)	Neurons	0.2478
	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	131 (11/6)	131 (11/6)	Neurons	0.0613
	Mean response layer2/3 (N_1) vs PE- (N_2)	Hierarchical bootstrap	870 (11/6)	131 (11/6)	Neurons	0.1381
	Mean response PE+ vs 0	Hierarchical bootstrap	131 (11/6)	-	Neurons	0.0055
	Mean response layer2/3 vs 0	Hierarchical bootstrap	870 (11/6)	-	Neurons	0.0990
	Mean response PE- vs 0	Hierarchical bootstrap	131 (11/6)	-	Neurons	0.2817

Table S1. Statistics

All information on statistical tests used in the manuscript are shown in Table S1. We used hierarchical bootstrap or a correlation coefficient for all comparisons.

Figure 5F	Correlation between activity and grating and activity and locomotion speed	-	870 (11/6)	-	Neurons	-
Figure 6D	Mean response ChrimsonR (N_1) vs Control stim (N_2)	Hierarchical bootstrap	664 (5/4)	664 (5/4)	Neurons	see Figure
Figure 6F	Mean response ChrimsonR (N_1) vs Control stim (N_2)	Hierarchical bootstrap	1308 (8/4)	1308 (8/4)	Neurons	see Figure
Figure 6G	Mean response Baz1a (N_1) vs Adamts2 (N_2)	Hierarchical bootstrap	664 (5/4)	1308 (8/4)	Neurons	$<10^{-4}$
	Mean response Baz1a (N_1) vs 0	Hierarchical bootstrap	664 (5/4)	-	Neurons	0.1274
	Mean response Adamts2 (N_1) vs 0	Hierarchical bootstrap	1308 (8/4)	-	Neurons	$<10^{-4}$
Figure 7B	Mean response Opto MM (N_1) vs Opto halt (N_2)	Hierarchical bootstrap	614 (6/4)	614 (6/4)	Neurons	see Figure
Figure 7C	Mean response Visuomotor MM	-	614 (6/4)	-	Neurons	-
Figure 8B	Mean response ChrimsonR (N_1) vs Control stim (N_2)	Hierarchical bootstrap	860 (9/4)	860 (9/4)	Neurons	see Figure
	Mean response ChrimsonR (N_1) vs no-ChrimsonR (N_2)	Hierarchical bootstrap	860 (9/4)	1394 (14/8)	Neurons	see Figure
	Mean response Control stim (N_1) vs no-ChrimsonR (N_2)	Hierarchical bootstrap	860 (9/4)	1394 (14/8)	Neurons	see Figure
Figure 8D	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	130 (9/4)	130 (9/4)	Neurons	see Figure
Figure 8E	Mean response PE+ (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	130 (9/4)	860 (9/4)	Neurons	0.2791
	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	130 (9/4)	130 (9/4)	Neurons	0.4569
	Mean response layer2/3 (N_1) vs PE- (N_2)	Hierarchical bootstrap	860 (9/4)	130 (9/4)	Neurons	0.4118
	Mean response PE+ vs 0	Hierarchical bootstrap	130 (9/4)	-	Neurons	$<10^{-4}$
	Mean response layer2/3 vs 0	Hierarchical bootstrap	860 (9/4)	-	Neurons	$<10^{-4}$
	Mean response PE- vs 0	Hierarchical bootstrap	130 (9/4)	-	Neurons	0.0303
Figure 8F	Correlation between activity and grating and activity and locomotion speed	-	860 (9/4)	-	Neurons	-
Figure S1A	Mean locomotion onset response closed loop (N_1) vs open loop (N_2)	Hierarchical bootstrap	6959 (69/36)	6959 (69/36)	Neurons	see Figure
Figure S1B	Mean locomotion onset response PE+ closed loop (N_1) vs open loop (N_2)	Hierarchical bootstrap	1045 (68/36)	1045 (68/36)	Neurons	see Figure
	Mean locomotion onset response PE- closed loop (N_1) vs open loop (N_2)	Hierarchical bootstrap	1045 (68/36)	1045 (68/36)	Neurons	see Figure
Figure S2A	Mean response Opto stim (N_1) vs Control stim (N_2)	Correlation	4262 (42/19)	4262 (42/19)	Neurons	r = 0.014 p = 0.3635
Figure S2B	Mean response Control stim (N_1) vs no-ChrimsonR (N_2)	Hierarchical bootstrap	4262 (42/19)	1394 (14/8)	Neurons	0.4995
Figure S3B	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	359 (35/17)	359 (35/17)	Neurons	see Figure
	Mean response PE+ vs 0	Hierarchical bootstrap	359 (35/17)	-	Neurons	see Figure
	Mean response PE- vs 0	Hierarchical bootstrap	359 (35/17)	-	Neurons	see Figure
Figure S3C	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	718 (35/17)	718 (36/17)	Neurons	see Figure
	Mean response PE+ vs 0	Hierarchical bootstrap	718 (35/17)	-	Neurons	see Figure
	Mean response PE- vs 0	Hierarchical bootstrap	718 (36/17)	-	Neurons	see Figure
Figure S3D	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	539 (35/17)	539 (35/17)	Neurons	see Figure
	Mean response PE+ (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	539 (35/17)	3591 (36/17)	Neurons	see Figure
	Mean response PE- (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	539 (35/17)	3591 (36/17)	Neurons	see Figure
Figure S4	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	144 (9/5)	144 (9/5)	Neurons	see Figure
	Mean response PE+ (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	144 (9/5)	960 (9/5)	Neurons	see Figure
	Mean response PE- (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	144 (9/5)	960 (9/5)	Neurons	see Figure
Figure S5A	Mean response Pref. Grat. (N_1) vs mean response Orth. Grat. (N_2)	Hierarchical bootstrap	3591 (36/17)	3591 (36/17)	Neurons	see Figure
Figure S5B	Mean response PE1+ (N_1) vs PE1- (N_2)	Hierarchical bootstrap	539 (36/17)	539 (35/17)	Neurons	see Figure
	Mean response PE1+ vs 0	Hierarchical bootstrap	539 (36/17)	-	Neurons	see Figure
	Mean response PE1- vs 0	Hierarchical bootstrap	539 (35/17)	-	Neurons	see Figure
	Mean response PE2+ (N_1) vs PE2- (N_2)	Hierarchical bootstrap	539 (36/17)	539 (35/17)	Neurons	see Figure
	Mean response PE2+ vs 0	Hierarchical bootstrap	539 (36/17)	-	Neurons	see Figure
	Mean response PE2- vs 0	Hierarchical bootstrap	539 (35/17)	-	Neurons	see Figure
Figure S5D	Mean response MM even (N_1) vs odd (N_2)	Correlation	3923 (39/17)	3923 (39/17)	Neurons	r = 0.589 $p < 10^{-4}$

Table S1. (continued)

	Mean response MM (N_1) vs inverted MM (N_2)	Correlation	754 (8/5)	754 (8/5)	Neurons	r = 0.403 $p < 10^{-4}$
Figure S5E	Mean response PE- vs 0	Hierarchical bootstrap	110 (8/5)	-	Neurons	see Figure
Figure S5F	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	110 (8/5)	110 (8/5)	Neurons	see Figure
	Mean response PE+ vs 0	Hierarchical bootstrap	110 (8/5)	-	Neurons	see Figure
	Mean response PE- vs 0	Hierarchical bootstrap	110 (8/5)	-	Neurons	see Figure
Figure S6B	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	150 (13/6)	150 (13/6)	Neurons	see Figure
	Mean response PE+ vs 0	Hierarchical bootstrap	150 (13/6)	-	Neurons	see Figure
	Mean response PE- vs 0	Hierarchical bootstrap	150 (13/6)	-	Neurons	see Figure
Figure S6C	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	300 (13/6)	300 (13/6)	Neurons	see Figure
	Mean response PE+ vs 0	Hierarchical bootstrap	300 (13/6)	-	Neurons	see Figure
	Mean response PE- vs 0	Hierarchical bootstrap	300 (13/6)	-	Neurons	see Figure
Figure S6D	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	225 (13/6)	225 (13/6)	Neurons	see Figure
	Mean response PE+ (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	225 (13/6)	1494 (13/6)	Neurons	see Figure
	Mean response PE- (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	225 (13/6)	1494 (13/6)	Neurons	see Figure
Figure S7B	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	88 (11/6)	88 (11/6)	Neurons	see Figure
	Mean response PE+ vs 0	Hierarchical bootstrap	88 (11/6)	-	Neurons	see Figure
	Mean response PE- vs 0	Hierarchical bootstrap	88 (11/6)	-	Neurons	see Figure
Figure S7C	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	173 (11/6)	173 (11/6)	Neurons	see Figure
	Mean response PE+ vs 0	Hierarchical bootstrap	173 (11/6)	-	Neurons	see Figure
	Mean response PE- vs 0	Hierarchical bootstrap	173 (11/6)	-	Neurons	see Figure
Figure S7D	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	131 (11/6)	131 (11/6)	Neurons	see Figure
	Mean response PE+ (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	131 (11/6)	870 (11/6)	Neurons	see Figure
	Mean response PE- (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	131 (11/6)	870 (11/6)	Neurons	see Figure
Figure S8A	Mean response layer 2/3 stationary (N_1) vs locomotion (N_2)	Hierarchical bootstrap	4241 (42/20)	4241 (42/20)	Neurons	see Figure
Figure S8B	Mean response layer 2/3 stationary (N_1) vs locomotion (N_2)	-	4241 (42/20)	4241 (42/20)	Neurons	-
Figure S8C	Left: Mean response sub.1 stationary (N_1) vs locomotion (N_2)	Hierarchical bootstrap	1276 (42/20)	1276 (42/20)	Neurons	see Figure
	Right: Mean response sub.2 stationary (N_1) vs locomotion (N_2)	Hierarchical bootstrap	2965 (42/20)	2965 (42/20)	Neurons	see Figure
Figure S8D	Mean response Tlx3 stationary (N_1) vs locomotion (N_2)	Hierarchical bootstrap	1740 (12/3)	1740 (12/3)	Neurons	see Figure
Figure S8E	Mean response Tlx3 stationary (N_1) vs locomotion (N_2)	-	1740 (12/3)	1740 (12/3)	Neurons	-
Figure S8F	Left: Mean response sub.1 stationary (N_1) vs locomotion (N_2)	Hierarchical bootstrap	398 (12/3)	398 (12/3)	Neurons	see Figure
	Right: Mean response sub.2 stationary (N_1) vs locomotion (N_2)	Hierarchical bootstrap	1342 (12/3)	1342 (12/3)	Neurons	see Figure
Figure S8G	Mean response layer 2/3 stationary (N_1) vs locomotion (N_2)	Hierarchical bootstrap	860 (9/4)	860 (9/4)	Neurons	see Figure
Figure S8H	Mean response layer 2/3 stationary (N_1) vs locomotion (N_2)	-	860 (9/4)	860 (9/4)	Neurons	-
Figure S8I	Left: Mean response sub.1 stationary (N_1) vs locomotion (N_2)	Hierarchical bootstrap	243 (9/4)	243 (9/4)	Neurons	see Figure
	Right: Mean response sub.2 stationary (N_1) vs locomotion (N_2)	Hierarchical bootstrap	617 (9/4)	617 (9/4)	Neurons	see Figure
Figure S9B	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	86 (9/4)	86 (9/4)	Neurons	see Figure
	Mean response PE+ vs 0	Hierarchical bootstrap	86 (9/4)	-	Neurons	see Figure
	Mean response PE- vs 0	Hierarchical bootstrap	86 (9/4)	-	Neurons	see Figure
Figure S9C	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	171 (9/4)	171 (9/4)	Neurons	see Figure
	Mean response PE+ vs 0	Hierarchical bootstrap	171 (9/4)	-	Neurons	see Figure
	Mean response PE- vs 0	Hierarchical bootstrap	171 (9/4)	-	Neurons	see Figure
Figure S9D	Mean response PE+ (N_1) vs PE- (N_2)	Hierarchical bootstrap	130 (9/4)	130 (9/4)	Neurons	see Figure
	Mean response PE+ (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	130 (9/4)	860 (9/4)	Neurons	see Figure
	Mean response PE- (N_1) vs layer 2/3 (N_2)	Hierarchical bootstrap	130 (9/4)	860 (9/4)	Neurons	see Figure

Table S1. (continued)

Number of mice	Genotype	Vector	Figures
10	Tlx3-Cre	AAV2/1-EF1α-GCaMP6f, AAV2/1-hSyn-DIO-ChrimsonR-tdTomato	**Figures 2, 3B-3F, S1, S2, S3B-S3D, S5A-S5D, S8A-C**
5	Tlx3-Cre	AAV2/1-EF1α-GCaMP6f, AAV2/1-hSyn-DIO-ChrimsonR-tdTomato	**Figures 2, 3B-3F, S1, S2, S3B-S3D, S5, S8A-S8C**
2	Tlx3-Cre	AAV2/1-EF1α-GCaMP6f, AAV2/1-hSyn-DIO-ChrimsonR-tdTomato	**Figures 3B-3F, 7B-7D, S2, S3B-S3D, S5A-S5D, S8A-S8C**
1	Tlx3-Cre	AAV2/1-EF1α-GCaMP6f, AAV2/1-hSyn-DIO-ChrimsonR-tdTomato	**Figures 3B-3C, 7B-7D, S2, S8A-S8C**
1	Tlx3-Cre	AAV2/1-EF1α-GCaMP6f, AAV2/1-hSyn-DIO-ChrimsonR-tdTomato	**Figures 7B-7D, S8A-C**
1	Tlx3-Cre	AAV2/1-EF1α-GCaMP6f, AAV2/1-hSyn-DIO-ChrimsonR-tdTomato	**Figures 3B-3C, S2, S8A-C**
3	Tlx3-Cre	AAV2/9-DIO-jGCaMP8s-P2A-ChrimsonR-ST	**Figure S8D-S8F**
6	Fezf2-CreERT2	AAV2/1-EF1α-GCaMP6f, AAV2/1-hSyn-DIO-ChrimsonR-tdTomato	**Figures 2, 4B-4F, S1, S6B-S6D**
6	Ntsr1-Cre	AAV2/1-EF1α-GCaMP6f, AAV2/1-hSyn-DIO-ChrimsonR-tdTomato	**Figures 2, 5B-5F, S1, S7B-S7D**
4	Scnn1a-Cre	AAV2/1-EF1α-GCaMP6f, AAV2/1-hSyn-DIO-ChrimsonR-tdTomato	**Figures 2, 8B-8F, S1, S8G-S8I, S9B-S9D**
5	C57BL/6	AAV2/1-EF1α-GCaMP6f	**Figures 2, 3B, 4B, 5B, 8B, S1, S2B, S4**
3	C57BL/6	AAV2/1-EF1α-GCaMP6f	**Figures 3B, 4B, 5B, 8B, S2B**
4	Tlx3-Cre x Ai148	AAV2/1-AP.Baz1a.1-CreOFF-ChrimsonR-mRuby2-ST	**Figures 6D, 6G**
4	Tlx3-Cre x Ai148	AAV2/1-AP.Adamts2.1-CreOFF-ChrimsonR-mRuby2-ST	**Figure 6F, 6G**

Table S2. Number of mice per experiment

Data availability

All raw data and software to generate all figures in this manuscript will be deposited to zenodo.org upon submission as the Version of Record. Software to control the two-photon microscope is available at *https://sourceforge.net/projects/iris-scanning/*.

Acknowledgements

We thank Ashena Gorgan Mohammadi, Manu Srinath Halvagal, and Friedemann Zenke for inspiration and discussion as well as comments on earlier versions of this manuscript. We thank Tingjia Lu for production of the AAV vectors, Jennifer Gröli for animal husbandry, and Z. Josh Huang for sharing Fezf2-CreERT2 mice. We have received generous donations of Dr. Pepper from Felix Widmer. We thank all the members of the Keller lab for discussion and support. This project has received funding from the Swiss National Science Foundation (GBK), the Novartis Research Foundation (GBK), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 865617) (GBK).

Additional information

Author contributions

AV designed and performed the experiments and analyzed the data. All authors wrote the manuscript.

Funding

Funder	Grant reference number	Author
EC | European Research Council (ERC)	https://doi.org/10.3030/865617	Georg B Keller
Novartis Foundation		Georg B Keller

Author ORCID iDs

Georg B Keller: https://orcid.org/0000-0002-1401-0117

References

Aizenbud I, Audette N, Auksztulewicz R, Basiński K, Bastos AM, Berry M, Canales-Johnson A, Choi H, Clopath C, Cohen U, *et al.* (2025) Neural mechanisms of predictive processing: a collaborative community experiment through the OpenScope program. *arXiv* https://doi.org/10.48550/arXiv.2504.09614

Anderson CT, Sheets PL, Kiritani T, Shepherd GMG (2010) Sublayer-specific microcircuits of corticospinal and corticostriatal neurons in motor cortex. *Nat. Neurosci* **13**:739-744 https://doi.org/10.1038/nn.2538 | PubMed

Assran M, Duval Q, Misra I, Bojanowski P, Vincent P, Rabbat M, LeCun Y, Ballas N (2023) Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture. *arXiv* https://doi.org/10.48550/arXiv.2301.08243

Attinger A, Wang B, Keller GB (2017) Visuomotor Coupling Shapes the Functional Development of Mouse Visual Cortex. *Cell* **169**:1291-1302.e14 https://doi.org/10.1016/j.cell.2017.05.023 | PubMed

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. *Neuron* **76**:695-711 https://doi.org/10.1016/j.neuron.2012.10.038 | PubMed

Bharioke A, Munz M, Brignall A, Kosche G, Eizinger MF, Ledergerber N, Hillier D, Gross-Scherf B, Conzelmann K.-K, Macé E, *et al.* (2022) General anesthesia globally synchronizes activity selectively in layer 5 cortical pyramidal neurons. *Neuron* **110**:2024-2040.e10 https://doi.org/10.1016/j.neuron.2022.03.032 | PubMed

Constantinople CM, Bruno RM (2013) Deep cortical layers are activated directly by thalamus. *Science* **340**:1591-1594 https://doi.org/10.1126/science.1236425 | PubMed

Daigle TL, Madisen L, Hage TA, Valley MT, Knoblich U, Larsen RS, Takeno MM, Huang L, Gu H, Larsen R, *et al.* (2018) A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. *Cell* **174**:465-480.e22 https://doi.org/10.1016/j.cell.2018.06.035 | PubMed

Douglas RJ, Martin K a.C, Whitteridge D (1989) A Canonical Microcircuit for Neocortex. *Neural Comput* **1**:480-488 https://doi.org/10.1162/neco.1989.1.4.480

Felleman DJ, Van Essen DC (1991) Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cereb. Cortex* **1**:1-47 https://doi.org/10.1093/cercor/1.1.1

Fiser A, Mahringer D, Oyibo HK, Petersen AV, Leinweber M, Keller GB (2016) Experience-dependent spatial expectations in mouse visual cortex. *Nat. Neurosci* **19**:1658-1664 https://doi.org/10.1038/nn.4385 | PubMed

Garner AR, Keller GB (2022) A cortical circuit for audio-visual predictions. *Nat. Neurosci* **25**:98-105 https://doi.org/10.1038/s41593-021-00974-7 | PubMed

Gerfen CR, Paletzki R, Heintz N (2013) GENSAT BAC Cre-Recombinase Driver Lines to Study the Functional Organization of Cerebral Cortical and Basal Ganglia Circuits. *Neuron* **80**:1368-1383 https://doi.org/10.1016/j.neuron.2013.10.016 | PubMed

Gong S, Doughty M, Harbaugh CR, Cummins A, Hatten ME, Heintz N, Gerfen CR (2007) Targeting Cre Recombinase to Specific Neuron Populations with Bacterial Artificial Chromosome Constructs. *J. Neurosci* **27**:9817-9823 https://doi.org/10.1523/JNEUROSCI.2707-07.2007 | PubMed

Grill J.-B, Strub F, Altché F, Tallec C, Richemond PH, Buchatskaya E, Doersch C, Pires BA, Guo ZD, Azar MG, *et al.* (2020) Bootstrap your own latent: A new approach to self-supervised Learning. *arXiv* https://doi.org/10.48550/arXiv.2006.07733

Hage TA, Bosma-Moody A, Baker CA, Kratz MB, Campagnola L, Jarsky T, Zeng H, Murphy GJ (2022) Synaptic connectivity to L2/3 of primary visual cortex measured by two-photon optogenetic stimulation. *eLife* **11**:e71103 https://doi.org/10.7554/eLife.71103 | PubMed

Heindorf M, Keller GB (2024) Antipsychotic drugs selectively decorrelate long-range interactions in deep cortical layers. *eLife* **12**:RP86805 https://doi.org/10.7554/eLife.86805 | PubMed

Jordan MI, Rumelhart DE (1992) Forward Models: Supervised Learning with a Distal Teacher. *Cogn. Sci* **16**:307-354 https://doi.org/10.1207/s15516709cog1603_1

Jordan R, Keller G (2023) The locus coeruleus broadcasts prediction errors across the cortex to promote sensorimotor plasticity eLife. *eLife* **12** https://doi.org/10.7554/eLife.85111 | PubMed

Jordan R, Keller GB (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 | PubMed

Kawato M (1999) Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol* **9**:718-27 https://doi.org/10.1016/s0959-4388(99)00028-8 | PubMed

Keller GB, Bonhoečer T, Hübener M (2012) Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* **74**:809-15 https://doi.org/10.1016/j.neuron.2012.03.040 | PubMed

Keller GB, Mrsic-Flogel TD (2018) Predictive Processing: A Canonical Cortical Computation. *Neuron* **100**:424-435 https://doi.org/10.1016/j.neuron.2018.10.003 | PubMed

Keller GB, Sterzer P (2024) Predictive Processing: A Circuit Approach to Psychosis. *Annu. Rev. Neurosci* **47**:85-101 https://doi.org/10.1146/annurev-neuro-100223-121214 | PubMed

Kim EJ, Juavinett AL, Kyubwa EM, Jacobs MW, Callaway EM (2015) Three Types of Cortical L5 Neurons that Dičer in Brain-Wide Connectivity and Function. *Neuron* **88**:1253-1267 https://doi.org/10.1016/j.neuron.2015.11.002 | PubMed

LeCun Y (2022) A Path Towards Autonomous Machine Intelligence Version 0.9. *OpenReview*

Leinweber M, Ward DR, Sobczak JM, Attinger A, Keller GB (2017) A Sensorimotor Circuit in Mouse Cortex for Visual Flow Predictions. *Neuron* **95**:1420-1432.e5 https://doi.org/10.1016/j.neuron.2017.08.036 | PubMed

Leinweber M, Zmarz P, Buchmann P, Argast P, Hübener M, Bonhoečer T, Keller GB (2014) Two-photon calcium imaging in mice navigating a virtual reality environment. *J. Vis. Exp. JoVE* e50885 https://doi.org/10.3791/50885 | PubMed

Lotter W, Kreiman G, Cox D (2016) Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning. In: 5th Int. Conf. Learn. Represent. ICLR 2017-Conf. Track Proc.

Madisen L, Zwingman TA, Sunkin SM, Oh SW, Zariwala HA, Gu H, Ng LL, Palmiter RD, Hawrylycz MJ, Jones AR, *et al.* (2010) A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci* **13**:133-140 https://doi.org/10.1038/nn.2467 | PubMed

Markov NT, Ercsey-Ravasz M, Van Essen DC, Knoblauch K, Toroczkai Z, Kennedy H (2013) Cortical high-density counterstream architectures. *Science* **342**:1238406 https://doi.org/10.1126/science.1238406 | PubMed

Matho KS, Huilgol D, Galbavy W, He M, Kim G, An X, Lu J, Wu P, Di Bella DJ, Shetty AS, *et al.* (2021) Genetic dissection of the glutamatergic neuron system in cerebral cortex. *Nature* **598**:182-187 https://doi.org/10.1038/s41586-021-03955-9 | PubMed

MICrONS Consortium (2025) Functional connectomics spanning multiple areas of mouse visual cortex. *Nature* **640**:435-447 https://doi.org/10.1038/s41586-025-08790-w | PubMed

Mohammadi AG, Halvagal MS, Zenke F (2025) Understanding cortical computation through the lens of joint-embedding predictive architectures. *bioRxiv* https://doi.org/10.1101/2025.11.25.690220

Morgenstern NA, Bourg J, Petreanu L (2016) Multilaminar networks of cortical neurons integrate common inputs from sensory thalamus. *Nat. Neurosci* **19**:1034-1040 https://doi.org/10.1038/nn.4339 | PubMed

Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol. Cybern* **66**:241-251 https://doi.org/10.1007/BF00198477 | PubMed

Nejad KK, Anastasiades P, Hertäg L, Costa RP (2025) Self-supervised predictive learning accounts for cortical layer-specificity. *Nat. Commun* **16**:6178 https://doi.org/10.1038/s41467-025-61399-5 | PubMed

Oliviers G, Tang M, Bogacz R (2025) Bidirectional predictive coding. *arXiv* https://doi.org/10.48550/arXiv.2505.23415

O'Toole SM, Oyibo HK, Keller GB (2023) Molecularly targetable cell types in mouse visual cortex have distinguishable prediction error responses. *Neuron* **111**:2918-2928.e8 https://doi.org/10.1016/j.neuron.2023.08.015 | PubMed

Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field ečects. *Nat. Neurosci* **2**:79-87 https://doi.org/10.1038/4580 | PubMed

Salvatori T, Pinchetti L, Millidge B, Song Y, Bao T, Bogacz R, Lukasiewicz T (2022) Learning on Arbitrary Graph Topologies via Predictive Coding. *arXiv* https://doi.org/10.48550/arXiv.2201.13180

Saravanan V, Berman GJ, Sober SJ (2020) Application of the hierarchical bootstrap to multi-level data in neuroscience. *Neurons Behav. Data Anal. Theory* **3** PubMed

Schneider-Mizell CM, Bodor AL, Brittain D, Buchanan J, Bumbarger DJ, Elabbady L, Gamlin C, Kapner D, Kinn S, Mahalingam G, *et al.* (2025) Inhibitory specificity from a connectomic census of mouse visual cortex. *Nature* **640**:448-458 https://doi.org/10.1038/s41586-024-07780-8

Solyga M, Keller GB (2025) Multimodal mismatch responses in mouse auditory cortex. *eLife* **13**:RP95398 https://doi.org/10.7554/eLife.95398 | PubMed

Spratling MW (2017) A review of predictive coding algorithms. *Brain Cogn* **112**:92-97 https://doi.org/10.1016/j.bandc.2015.11.003 | PubMed

Straka Z, Svoboda T, Hočmann M (2020) PreCNet: Next Frame Video Prediction Based on Predictive Coding. *arXiv* https://doi.org/10.48550/arxiv.2004.14878

Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, *et al.* (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci* **19**:335-346 https://doi.org/10.1038/nn.4216 | PubMed

Vesuna S, Kauvar IV, Richman E, Gore F, Oskotsky T, Sava-Segal C, Luo L, Malenka RC, Henderson JM, Nuyujukian P, *et al.* (2020) Deep posteromedial cortical rhythm in dissociation. *Nature* **586**:87-94 https://doi.org/10.1038/s41586-020-2731-9 | PubMed

Whittington JCR, Bogacz R (2017) An approximation of the error backpropagation algorithm in a predictive coding network with local hebbian synaptic plasticity. *Neural Computation* **29**:1229-1262 https://doi.org/10.1162/NECO_a_00949

Widmer FC, O'Toole SM, Keller GB (2022) NMDA receptors in visual cortex are necessary for normal visuomotor integration and skill learning. *eLife* **11**:e71476 https://doi.org/10.7554/eLife.71476 | PubMed

Zmarz P, Keller GB (2016) Mismatch Receptive Fields in Mouse Visual Cortex. *Neuron* **92**:766-772 https://doi.org/10.1016/j.neuron.2016.09.057 | PubMed

Peer reviews

Reviewer #1 (Public review):

Vasilevskaya and Keller test different models of cortical function through the lens of predictive processing, a powerful framework for the brain to learn and predict the statistics of the world via generative internal models. The authors use a clever combination of behavioral perturbations in closed-loop and open-loop visuomotor virtual reality assays, a paradigm the Keller lab pioneered and used effectively in the past decade, in conjunction with two photon imaging of neuronal calcium responses and targeted optogenetic perturbations of activity. They specifically put to test proposed hierarchical vs. non-hierarchical circuit implementations of predictive processing by analyzing the logic of inter-lamina interactions (superficial vs. deep; L2/3 vs. L5/6).

The authors conclude that both versions of predictive processing architectures they analyze are likely invalid and instead formulate an alternative novel model of cortical function based on a recently developed machine learning algorithm for self-supervised learning (joint embeddings of predictive architectures, JEPA) and its further refinements. JEPA borrows elements from predictive processing engaging two encoder networks and training the output of one network to predict the output of the other. In their new model of cortical computations, prediction errors neurons in L2/3 compare the deep layers (L5/6) activity, which is taken as a teaching signal, to a local, L2/3 prediction of this latent representation.

Specifically, the authors build on their previous work and reports from other groups that different sets of L2/3 neurons compute positive prediction errors (fire when sensory stimuli appear unexpectedly with respect to the movements of the animal; e.g., grating onsets in the absence of locomotion) and respectively negative prediction errors (fire when sensory stimuli are absent, while the brain expected them to be present; e.g. mice locomote but visual flow is suddenly halted - visuomotor mismatches). These L2/3 positive and negative prediction error neurons exchange messages with neurons in the deeper cortical layers that, the authors propose, build an internal representation (R) of the sensory stimuli given the animals' movements.

In the hierarchical model, internal representation neurons (R) are supposed to act as a teaching signal for both types of prediction error neurons; the output of the positive prediction error neurons is assumed to suppress activity of R such that the error between the

teaching signal and the prediction is minimized; similarly, in the non-hierarchical version, R serves as a prediction for the prediction error neurons, and in turn it receives excitatory drive from the positive prediction error neurons and negative input from the negative prediction error neurons.

The authors find that the functional impact of L5 neurons to L2/3 neurons is not compatible with the non-hierarchical architecture they and other groups proposed, but rather in accordance with the hierarchical model. At the same time, the functional impact of L2/3 neurons (positive vs. negative prediction error neurons) on L5 neurons (internal representation) appears not compatible with the hierarchical model, but rather in accordance with the non-hierarchical implementation.

They further hypothesize that L2/3 prediction error neurons don't use sensory input, but rather the L5 activity as a teaching signal, and test it using perturbations (halts) of optogenetic stimulation of L5 neurons coupled with locomotion (Fig.7).

All in all, the question is topical, and the new model addresses a decades-long quest to develop a unifying model of cortical function. The findings reported here transform our understanding of cortical computations, opening new exciting avenues for future investigation. The experimental design and execution are rigorous; the arguments are clearly laid out (in spite of ample potential for confusion given the numerous loops and sign flips). These include a discussion of why the non-hierarchical model proposed by the same group does not hold, as well as potential caveats in interpreting the results and novel testable proposed experiments emerging from the JEPA-like model.

Comments on revised version

I commend the authors for nicely provided answers and addressing my concerns and clarifying their points on the relationship of their findings to JEPA and current state of understanding.

In particular for Q5 -I meant if there is also a positive correlation between optomotor mismatch response and visuomotor mismatch response when looking only at neurons that the authors identify as PE-?

The authors provided the answer on point.

Overall, I think this is a fundamental study and the strength of evidence for the claims is exceptional as an exemplary use of existing approaches.

https://doi.org/10.7554/eLife.110827.2.sa3

Reviewer #2 (Public review):

This manuscript reveals functional connectivity of two different classed of cortical neurons that respond in opposite ways to mismatches between sensory and top-down inputs. These data are very valuable because different theories of information processing in the cortex make different predictions on the patterns of connectivity of these neurons. Therefore, these data strongly constrain possible theories of cortical processing.

Comments on revised version.

I thank the Authors for answering my questions and updating the manuscript.

Congratulations on this important work!

https://doi.org/10.7554/eLife.110827.2.sa2

Reviewer #3 (Public review):

Vasilevskaya and Keller set out to experimentally distinguish between two variants of predictive processing: a hierarchical and a non-hierarchical variant. The hierarchical variant assumes a hierarchical organization in which internal representation neurons (believed to be a subset of layer 5 excitatory neurons) serve as a source of a teaching signal for local prediction error neurons as well as for the next higher level of the hierarchy, while simultaneously providing prediction signals to the preceding lower level. In contrast, the non-hierarchical variant posits that these layer 5 internal representation neurons provide local predictions to layer 2/3 prediction error neurons.

The interaction between internal representation neurons and prediction error neurons differs fundamentally between the two variants. In the hierarchical variant, internal representation neurons excite positive prediction error neurons and inhibit negative prediction error neurons, while at the same time being inhibited by positive prediction error neurons and excited by negative prediction error neurons. In the non-hierarchical variant, this pattern of connectivity is reversed.

This work is very exciting, timely, and carefully executed. The authors functionally, and later molecularly, identify layer 2/3 prediction error neurons in V1 and probe their interactions with genetically defined neuron types in cortical layers 5 and 6 using optogenetics. They demonstrate that the functional influence of putative prediction error neurons in layer 2/3 onto layer 5 is incompatible with the hierarchical variant, whereas the influence of layer 5 onto putative prediction error neurons in layer 2/3 is incompatible with the non-hierarchical variant. They then test an alternative hypothesis, in which layer 2/3 responses resemble prediction errors with respect to perturbations of artificial layer 5 activity patterns. To investigate this, they designed an experiment in which optogenetic activation of L5 IT neurons was closed-loop coupled to the mouse's locomotion speed in the absence of visual feedback, allowing them to probe the causal influence of L5 activity on layer 2/3 responses.

Finally, the authors hypothesize that their data are more consistent with a joint embedding predictive architecture (JEPA) and outline experimentally testable predictions arising from this framework.

While the work is overall convincing and provides important insights into the circuit-level implementation of predictive processing, I think the connection to JEPA networks would benefit from a more in-depth discussion of its relationship to recently proposed and implemented models. Below, I address the specific points raised by the authors (flanked by ' ... ' to make the author's statements stand out), in particular in relation to the model proposed by Nejad et al. (2025):

- 'The two proposals indeed share similarities in assuming that bottom-up input for both L2/3 and L5 arrives from thalamus, and that representations formed in L2/3 are used for predicting the activity of L5. However, there are a few important differences between the JEPA implementation proposal formulated here and the Nejad et al. model.

(1) There is no proposed mapping of computations in the Nejad et al. model onto different JEPA networks. We assume that the suggested mapping would be L4 and L5 as encoder networks, and L2/3 as a predictor network? In that case, it is different to our proposal, in which L2/3 is part of the encoder network.'

I think there is some confusion here. In Nejad et al., both L2/3 and L5 function as encoder networks. Each receives sensory input (with L2/3 receiving this input delayed via L4) and computes a latent representation of that input, denoted z_{L2/3} and z_{L5}, respectively. The prediction is obtained by comparing the output of L2/3 with L5 latent representations (via W_{L2/3->L5} * z_{L2/3}). In other words, L5 is the target in the learning objective.

Although, they did not explicitly state the mapping with JEPA, their model has the same fundamental property - predictive learning happens in the latent space. Also, this appears very similar to the roles assigned to L2/3 and L5 in your Figure 9B. The fact that you made this more explicit and the new data included, is in my view, a very interesting contribution. However, from an architectural perspective, it appears that your proposal and the model of Nejad et al. are conceptually very similar, and I do not see a substantial difference between the two. This should be made more clear in the Discussion.

- '2. Our proposal contains explicit prediction error neuron cell types within L2/3, while prediction errors in Nejad et al. are encoded in the gradients, and the layer origin of these signals is hypothesized to be L5 ('the learning-driving error signal originates in L5'). Hence, also the role of L5-L2/3 connection is distinct between the two proposals. In Nejad et al. this connection serves the role of error propagation and update for predictions in L2/3, while in our proposal this connection contains teaching signal (target representations) that are compared to predictions within L2/3. Similarly, the functional role of L2/3-L5 connection is also different, since in Nejad et al, it is supposed to carry predictions of L5 activity, whereas in our proposal we expect it to drive plasticity in L5 encoder.'

Indeed, in Nejad et al., the layer-dependent mismatch responses are modeled as gradients with respect to neuronal activity, and the model does not explicitly include prediction error neurons. However, this appears to be a modeling choice rather than a fundamental aspect of the proposal, and it does not preclude an implementation with explicit prediction error neurons. In fact, the authors explicitly acknowledge this possibility in the Discussion:

"The second approach would be to recast our model within a predictive coding framework... Predictive coding jointly optimizes both model parameters and neuronal activities, which could naturally lead to prediction errors observable in the activity of both L2/3 and L5 neurons. Note that these two views are not mutually exclusive."

While Nejad et al. hypothesize that the learning-driving error signals (that are distinct from their mismatch responses) originate in L5, the abstract loss function itself does not uniquely specify where the underlying comparison between the predicted representation (W_{L2/3 -> L5} z_{L2/3}) and the target representation (z_{L5}) must be implemented. The proposed biological implementation places this computation in L5, but from my understanding, the computational objective itself does not require this specific localization.

That said, I agree that your proposed model introduces a genuine difference. The functional roles assigned to the vertical projections are effectively reversed: in Nejad et al., the L2/3->L5 projection carries the prediction, whereas the L5->L2/3 projection conveys the error/gradient. In your architecture, by contrast, the L5->L2/3 projection carries the teaching signal (target). This is, in my view, a real and testable interpretational divergence that is worth stating clearly.

Therefore, I think the novelty lies less in the computational architecture itself and more in committing to a particular biological implementation-one that adds cell-type-specific detail to an implementation that Nejad et al. hypothesised as being compatible with their framework.

- '3. The difference outlined above also makes it evident that the two proposals should differ in how deep and superficial layers are expected to influence the activity of one another. Indeed, the proposal in Nejad et al. is based on the cortical column idea, and according to eq.

2 and 3 in the Methods, activity in L5 is a function of activity in L2/3, while activity in L2/3 is not a function of activity in L5. Our proposal is based on idea of layers forming parallel networks, where horizontal communication is the dominant mode of cortico-cortical interactions, and activity in deep layers serve as a teaching signal for L2/3. In our case, we expect the opposite - that activity in L2/3 depends on activity of L5, while activity of L5 is not immediately dependent on activity of L2/3 (only via plasticity route). This led us to propose one of direct tests for our framework - silencing L2/3 in a familiar setting should result in no immediate changes to L5 activity and behavior of the animal.'

My reading of Nejad et al. is consistent with your interpretation of the equations. Specifically, Eq. 3 makes L5 activity depend on L2/3 activity (albeit weakly, with a = 0.3), whereas Eq. 2 contains no L5 term, so L2/3 activity does not depend directly on L5 activity. In that model, the L5->L2/3 pathway carries the learning gradient rather than contributing to the activity dynamics. By contrast, in your proposed model, L2/3 activity depends on L5 activity, whereas L5 activity is not immediately dependent on L2/3 activity (except indirectly through learning/plasticity). So, you state that "silencing L2/3 in a familiar setting should result in no immediate changes in L5 activity.

[...].

However, I am unsure how to reconcile this prediction with the results shown in Fig. 6. If I understand the figure correctly, optogenetic activation of Rrad-positive (positive prediction error) L2/3 neurons produces a small increase in L5 activity, whereas activation of Adamts2-positive (negative prediction error) L2/3 neurons produces a decrease in L5 activity. Although these experiments involve activation rather than silencing, they nevertheless suggest that perturbing L2/3 activity can have an immediate effect on L5 activity. Could you clarify how this is consistent with the proposed model? In other words, what aspect of the proposed circuitry makes activation effective while silencing is predicted to have no immediate consequence?

For comparison, Nejad et al. performed a related perturbation analysis in Fig. S15 by scaling the output of L2/3 neurons exhibiting positive mismatch signals (defined through the activity gradients), which increased L5 activity, whereas scaling neurons with negative mismatch signals produced the opposite effect. I am not entirely sure how directly these simulations map onto the experiments shown in your Fig. 6, since the Nejad simulations were performed during mismatch conditions, if I have understood them correctly.

- '4. The proposal in Nejad et al. relies on input reconstruction or variance maximization within the L5 autoencoder network to avoid collapse. Instead, our proposal has no reconstruction objective.'

Nejad et al. only require two encoders (like in JEPA), how these two are learnt can be done in several ways. While Nejad et al. focus on using a reconstruction loss to learn the L5 target, they also show that it works equally well with non-reconstruction objectives. Therefore, I do not think the presence or absence of a reconstruction objective constitutes a fundamental distinction between the two proposals.

As your current work presents a conceptual architecture rather than a fully implemented learning algorithm (in a model), the mechanism that would prevent representational collapse has not yet been defined. From my understanding, every joint-embedding approach must address this issue, whether through reconstruction, variance/covariance regularization, stop-gradient or EMA mechanisms, or other approaches. Thus, the absence of a reconstruction objective (or another anti-collapse mechanism) is not, in itself, a distinguishing feature of the proposed architecture, but rather an as-yet unspecified design choice within the learning objective.

- '5. Lastly, there is time-delay between inputs to L5 and L2/3 that is proposed in Nejad et al., while this is not something inherent to our proposal.'

I agree that the temporal delay introduced by L4 is a key component of the Nejad et al. model and is currently absent from your proposal. However, I would expect temporal delays to emerge naturally in your framework as well, given the multisynaptic and highly parallel organization of cortical circuits. More generally, implementing predictive learning over time (as in JEPA) requires comparing representations at times t and t+1, which seems to require some form of temporal delay. How else would you suggest this is done?

In general, I think the manuscript would benefit from a clearer discussion of its relationship to the model proposed by Nejad et al. (perhaps following the discussion above), including both the shared conceptual claims and the aspects that genuinely differ between the two frameworks. At present, some of the claims are presented as novel, although at least some of these core ideas have already been proposed in Nejad et al.

For example, the authors state: "Thus, we propose that layer 2/3 functions to predict layer 5 activity, not sensory input per se, hence making predictions in the internal representation space, not input space." This appears to be exactly what Nejad et al. proposed as discussed above - in their model L2/3 predicts L5 activity (purely in the latent space), as they state in the abstract.

That said, there are some interesting differences, and I think the community would greatly benefit from making these clear, including the roles assigned to interlaminar connections and the interpretation of the signals carried by these pathways. These differences are interesting and potentially testable, and I think the manuscript would be strengthened by explicitly distinguishing which aspects are in line with the ideas already present in Nejad et al. and which aspects represent new contributions.

https://doi.org/10.7554/eLife.110827.2.sa1

Author response:

The following is the authors' response to the original reviews.

We thank you for the time you took to review our work and for your feedback! We have performed the following additional analyses:

(1) We analyzed the optomotor mismatch response as a function of time spent in the optogenetic closed-loop session.

(2) We show the correlation between optomotor mismatch response and visuomotor mismatch response for functionally identified PE neurons.

The main changes to the manuscript are:

(3) Clarification of our terminology (moved and refined definition of the teaching signals).

(4) A new figure panel summarizing the functional influence patterns we identified.

(5) Clarification of our argumentation for the JEPA-inspired proposal.

All comments are addressed individually in the following.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public review):***

Vasilevskaya and Keller test different models of cortical function through the lens of predictive processing, a powerful framework for the brain to learn and predict the statistics of the world via generative internal models. The authors use a clever combination of behavioral perturbations in closedloop and open-loop visuomotor virtual reality assays, a paradigm the Keller lab pioneered and used effectively in the past decade, in conjunction with two-photon imaging of neuronal calcium responses and targeted optogenetic perturbations of activity. They specifically put to test proposed hierarchical vs. non-hierarchical circuit implementations of predictive processing by analyzing the logic of inter-lamina interactions (superficial vs. deep; L2/3 vs. L5/6).

The authors conclude that both versions of predictive processing architectures they analyze are likely invalid, and instead formulate an alternative novel model of cortical function based on a recently developed machine learning algorithm for self-supervised learning (joint embeddings of predictive architectures, JEPA) and its further refinements. JEPA borrows elements from predictive processing, engaging two encoder networks and training the output of one network to predict the output of the other. In their new model of cortical computations, prediction error neurons in L2/3 compare the deep layers (L5/6) activity, which is taken as a teaching signal, to a local, L2/3 prediction of this latent representation.

Specifically, the authors build on their previous work and reports from other groups that different sets of L2/3 neurons compute positive prediction errors (fire when sensory stimuli appear unexpectedly with respect to the movements of the animal; e.g., grating onsets in the absence of locomotion) and respectively negative prediction errors (fire when sensory stimuli are absent, while the brain expected them to be present; e.g. mice locomote but visual flow is suddenly halted - visuomotor mismatches). These L2/3 positive and negative prediction error neurons exchange messages with neurons in the deeper cortical layers that, the authors propose, build an internal representation (R) of the sensory stimuli given the animals' movements.

In the hierarchical model, internal representation neurons (R) are supposed to act as a teaching signal for both types of prediction error neurons; the output of the positive prediction error neurons is assumed to suppress activity of R such that the error between the teaching signal and the prediction is minimized; similarly, in the non-hierarchical version, R serves as a prediction for the prediction error neurons, and in turn it receives excitatory drive from the positive prediction error neurons and negative input from the negative prediction error neurons.

The authors find that the functional impact of L5 neurons on L2/3 neurons is not compatible with the non-hierarchical architecture they and other groups proposed, but rather in accordance with the hierarchical model. At the same time, the functional impact of L2/3 neurons (positive vs. negative prediction error neurons) on L5 neurons (internal representation) appears not compatible with the hierarchical model, but rather in accordance with the non-hierarchical implementation.

They further hypothesize that L2/3 prediction error neurons don't use sensory input, but rather the L5 activity as a teaching signal, and test it using perturbations (halts) of optogenetic stimulation of L5 neurons coupled with locomotion (Figure 7).

All in all, the question is topical, and the new model addresses a decades-long quest to develop a unifying model of cortical function. The findings reported here transform our understanding of cortical computations, opening new, exciting avenues for future investigation. The experimental design and execution are rigorous; the arguments are clearly laid out (in spite of ample potential for confusion given the numerous loops and sign flips). These include a discussion of why the non-hierarchical model proposed by the

> *same group does not hold, as well as potential caveats in interpreting the results and novel testable proposed experiments emerging from the JEPA-like model.*
>
> *I have several questions about the interpretations of some of the claims and suggestions for potential additional experiments and analyses.*

We thank the reviewer for their comments. We address them below.

> *(1) Some of the pieces of the puzzle remain to be identified and demonstrated: the existence of internal representation neurons in L2/3 and ascertaining that the L5/6 neurons analyzed function indeed as internal representation neurons. The authors find that stimulation of L2/3 positive prediction error neurons enhances activity of L5 neurons...If L5 neurons hold a latent representation that serves as a teaching signal for L2/3 neurons (as the authors posit), wouldn't one expect that the input they receive from the positive prediction neurons be suppressive, such that the error is further minimized?*

Not necessarily - this depends on the model one has for how cortex works. In the hierarchical predictive processing model, PE+ neurons are expected to suppress the local internal representation neurons in L5. Our data, however, are not consistent with this model. This is one of the key arguments we build the idea on that JEPA is a better model for cortex than hierarchical predictive processing. In JEPA we would not expect the L2/3 prediction error to update L5 directly, but instead update the prediction of L5 activity (the source of this signal remains to be identified; see our speculations on the origin of prediction signals in the comments to question 3), and drive plasticity in the local L5 encoder network. That something acts like a teaching signal for the L2/3 comparator does not, by itself, determine how the outcome of the comparison influences the source of the teaching signal.

> *(2) Do the authors envision any specific differences between the representations of the two encoder networks posited to exist in L2/3 and L5 in the JEPA-like implementation? Are they synchronous/offset in their temporal representations, or any other features?*

Given theoretical work (Mohammadi et al., 2025), one would expect to find differences in learning rates between the predictor and the encoder networks. Assuming the predictor network is also implemented in L2/3, we would expect to see faster learning rates in L2/3 compared to L5. Implementations inspired by related theoretical works also predict differences in learning rate between the two encoder networks (Grill et al., 2020), again with higher learning rates expected in L2/3 compared to L5. Beyond that, however, we are not aware of any experimentally observable differences one might expect to find. We are hoping the computational community will remedy this soon.

> *(3) Where is the prediction coming from onto L2/3 neurons? Is it emerging locally in L2/3 from the putative internal representation neurons, or is it long-range - as work from the authors previously proposed? Or a mix of both?*

We expect the predictions to come from long-range inputs. In classical (hierarchical) JEPA one would expect these to be the lateral communication within L2/3. In a non-hierarchical implementation that is capable of operating on arbitrary graphs (as would be necessary for it to work in cortex), we suspect that the L2/3 network will use both long-range L2/3 and long-range L5 input. In JEPA terminology: the encoder A network uses information from non-local sources of both networks to predict local activity of encoder B network (assuming that information is statistically useful in predicting that activity).

> *(4) What is the role of the indiscriminate L4 input that appears to enhance activity of both positive and negative prediction error neurons in L2/3?*

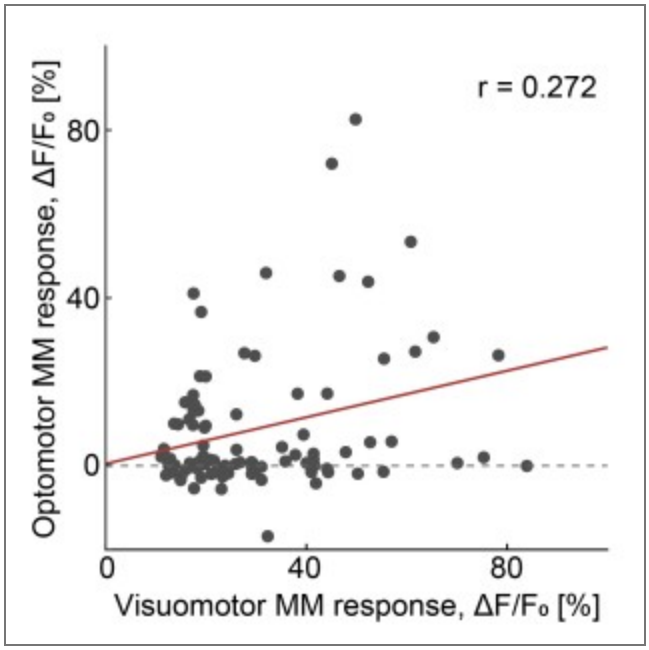
The short answer is, we don’t know. We would have indeed expected to find some asymmetry of influence. There are a few options: A) We might be failing to activate a specific interneuron

that mediates feedforward inhibition on this pathway by the artificial stimulation of Scnn1a neurons. B) We know that Scnn1a neurons are only a subset of L4 neurons – there are other populations of L4 neurons that might exhibit the opposing influence. C) The Scnn1a population is more than 1 layer of a JEPA away from the comparator and not part of the predictor that forms the actual representation compared against L5 (e.g. L4 could provide an input to L2/3 internal representation neurons, and thus only indirectly influence L2/3 PE neurons). D) The JEPA analogy is wrong.

> *(5) Does Figure 7D change in a meaningful manner if the authors plot the correlation between optomotor mismatch response and visuomotor mismatch response specifically for the negative prediction error neurons in L2/3 (Adamts-2) rather than for all L2/3 cells sampled?*

We might be misunderstanding. If the reviewer means genetically identified negative prediction error neurons (Adamts2), we do not have the data to address this question, as we did not perform any recordings of molecularly defined Adamts2 population in L2/3. If the reviewer means functionally identified negative prediction neurons, this would be the two rightmost data points in Figure 7D (the x-axis in this panel is the visuomotor mismatch response strength we use to functionally identify PE- neurons). These two data points on the right-hand side of the plot correspond exactly to what we classify as negative prediction error neurons throughout the rest of the manuscript (15% of the most responsive neurons to visuomotor mismatch).

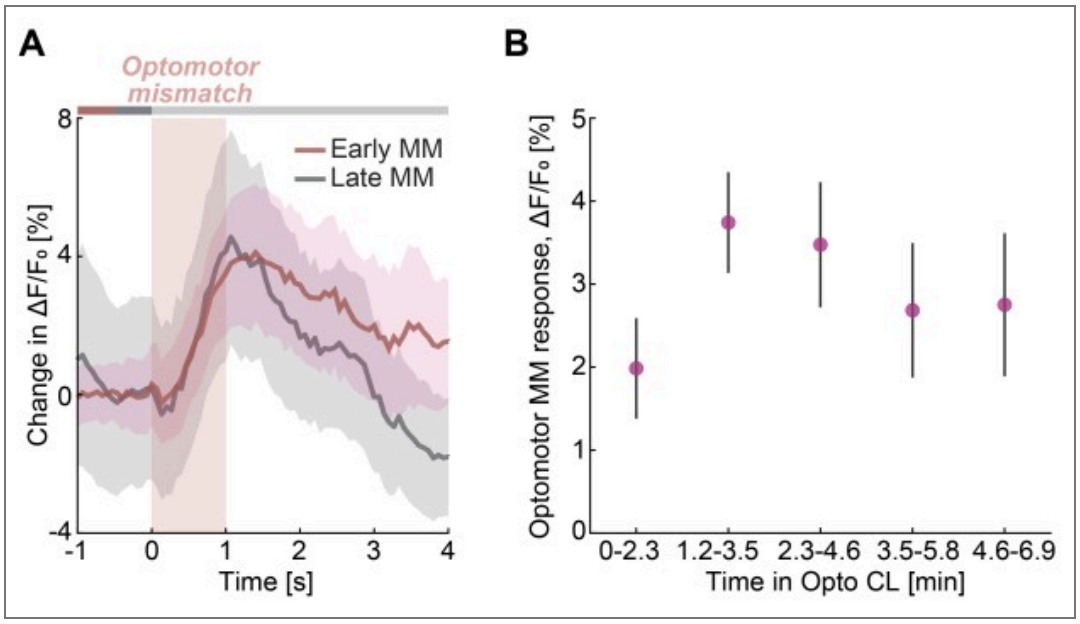
Does the reviewer mean, is there also a positive correlation between optomotor mismatch response and visuomotor mismatch response when looking only at neurons that we identify as PE-? If so, the answer is yes (Author response image 1). Interestingly, while there is a positive correlation, there is also nonuniformity in response patterns. We think that this is expected, given that our visuomotor coupling paradigm captures only a very small subspace of stimuli that prediction error neurons are tuned to. Bulk stimulation of L5, in contrast, might work better to separate all PE neurons. In other words, we speculate that L5 stimulation is a better predictor of the functional role of an L2/3 neuron.



Author response image 1. Optomotor mismatch response as a function of visuomotor mismatch response for neurons that are functionally classified as PE. Red line shows a linear fit estimated with a bootstrap approach.

> *(6) Do the optomotor mismatch responses in L2/3 neurons depend on how long the closed-loop coupling of optogenetic stimulation of Tlx3 L5 neurons and locomotion speed has been in place for?*

No, not that we can measure. We performed an analysis in which we split the optogenetic closed-loop session into two equal parts, early and late. We then quantified the average optomotor mismatch response independently for early and late parts of the session (Author response image 2A). Based on this quantification, we find no evidence of a change as a function of time in the closed-loop session. We also performed a sliding window analysis on a shorter timescale. While it did look like the responses may be smaller in the first few minutes, we do not have sufficient data to address this. None of the differences in response size were significant (Author response image 2B).



Author response image 2. Optomotor mismatch response as a function of experience with artificial closed-loop coupling. (A) Mean L2/3 population response to optomotor mismatch in the first half (Early MM) and in the second half (Late MM) of the optogenetic closed-loop session. (B) Mean L2/3 population response to optomotor mismatch as a function of time in the optogenetic closed-loop session. Error bars indicate SEM. Differences between the mean values were estimated by hierarchical bootstrap and are not significant.

> ***Reviewer #2 (Public review):***
>
> *This manuscript reveals the functional connectivity of two different classes of cortical neurons that respond in opposite ways to mismatches between sensory and top-down inputs. These data are very valuable because different theories of information processing in the cortex make different predictions on the patterns of connectivity of these neurons. Therefore, these data strongly constrain possible theories of cortical processing.*

We thank the reviewer for their comments. We address them below.

> *General comments:*
>
> *(1) The methods of statistical testing are insufficiently described. I did not understand the description in lines 1105-1119. The authors should provide sufficient details so the reader can reproduce their analyses. For example, it may be helpful to provide specific details of the testing procedure for one of the comparisons (e.g. the first comparison in Table S1).*

We assume the reviewer is not familiar with hierarchical bootstrapping in general. If so, the explanation below would summarize the procedure. This is the procedure with particular emphasis on its application to neuroscience data is described in the paper we reference in that part of the methods (Saravanan et al., 2020). Given that the analysis has become relatively standard (and is described in the reference provided), we think it might be an overkill to add the full explanation below to the manuscript. In addition to the general procedure, there are only 2 pieces of information relevant to fully reconstructing the analysis:

(1) What are the “levels” (mice, recording sites, neurons, trials)?

(2) What is the number of bootstrap samples used.

Thus, we think all information is already provided in the methods. We now also explicitly added the levels when describing the nested structure of the data in the manuscript to

increase clarity.

Hierarchical bootstrap analysis:

Our data are naturally nested: multiple neurons are recorded within a single mouse, and multiple mice are tested within an experimental group. Standard bootstrapping (sampling with replacement from the entire pool of neurons) fails because it assumes all observations are independent. In reality, neurons from the same mouse are more similar to each other than to neurons from a different mouse. The hierarchical bootstrap (or multi-level bootstrap) preserves this nested structure, ensuring your confidence intervals are not artificially narrow due to pseudoreplication.

To illustrate the problem, assume you have 100 neurons from Mouse A and 10 neurons from Mouse B, a simple bootstrap will be heavily biased toward Mouse A. Furthermore, the simple bootstrap ignores the fact that the true variance in your population comes from two sources:

(1) Between-mouse variance (differences in surgery, genetics, or behavior).

(2) Within-mouse variance (differences in tuning or activity between individual cells).

Hierarchical bootstrap addresses this problem, and is implemented as follows: To estimate the mean response while accounting for different sample sizes per mouse, a two-level resampling scheme is used

(1) Resample the higher level (Mice)

First, you account for the variability between animals.

Suppose you have N mice in total.

Randomly draw N mice with replacement from your original pool.

Note: Because this is with replacement, a single mouse's data might be included multiple times in one bootstrap iteration, while another mouse might be left out entirely.

(2) Resample the lower level (Neurons)

For each mouse selected in Step 1, you must now account for the variability within that specific animal.

Look at the number of neurons actually recorded from that mouse (let's call it k_i).

Randomly draw k_i neurons with replacement from that mouse's specific pool of recorded cells.

This step is crucial: you always resample the same number of neurons that were originally recorded for that specific mouse. This maintains the "weight" or "influence" that animal had in the original dataset.

(3) Calculate the resampled statistic

Calculate the mean of all neurons collected in this "bootstrap sample".

(4) Iterate

Repeat Steps 1–3 many times (typically B = 1,000 or 10,000 iterations).

The distribution of these B bootstrap means represents your sampling distribution and is used to calculate confidence intervals and p-values.

> *(2) The authors should clarify how the problem of multiple comparisons was addressed for comparisons performed in multiple moments of time, where significance is indicated by a black bar (e.g. in Figure 2F).*

There is no family-wise error correction in cases of comparing response time courses implemented in our analysis. If the reviewer has a good suggestion for how to implement family wise error correction, we would be happy to implement it. We are not aware of anything that is better than what we currently do (the time bin-wise comparison). To briefly explain the problem: In most neuroscience papers, response curves are compared by choosing a time window (e.g. 0.5 to 1 s following a trigger) and calculating mean values of the curves in these windows. This hides the problem of multiple comparisons that arises from the fact that the experimenter is free to choose a response window used for analysis. Note, this also creates a strong incentive to "optimize" choice of an analysis window – a part of the analysis that a reader is typically completely blind to. We could of course also choose an analysis window that sounds reasonable and yields significant differences for our analyses. However, to provide a more unbiased image of the data, we have come to do bin-wise comparisons with a fixed p-value (typically 0.05). This is used in all of our papers at the moment. Given that samples from neighboring timepoints are correlated via a combination of actual responses and a subset of noise sources, the samples are not independent. We now also implemented a correction for spurious positive values by requiring at least 2 neighboring bins to have a p-value below 0.05 to be shown. If the samples were independent, this would mean a false positive rate of 0.0025. Given that they are not, this is a lower bound only. Additionally, any family-wise error correction would be a function of the number of time bins we show in the plot. This would mean that our choice of the time window shown in a plot (-1s to +4s, or +5s, etc.) would change the p-value we consider significant. Thus, there is no explicit family-wise error correction, and we have come to the conclusion that the bin-wise comparison with a fixed p-value is the most unbiased representation of the data we can provide. The alternative would be to additionally plot z-scores or the p-values as a function of time, but in our experience these types of plots are even harder to read for readers not used to it.

> *(3) It would be helpful to add a figure in the Discussion summarising the functional connectivity suggested by all experiments.*

We now added a panel that summarizes the functional connectivity observed in our experiments to Figure S10.

> *(4) Throughout the manuscript, the authors use the term "teaching signals", but I am unclear what they mean by it: after reading the definition in lines 45-46, I thought that they corresponded to values (as they are compared to sensory signals). Later (428-430), the text suggests that they correspond to error neurons. But then lines 605-607 say it is not an error signal. The authors should define teaching signals very precisely or remove this term.*

The formal definition of the teaching signal was in footnote 1 of the manuscript. We assume the reviewer may have missed this. We now moved this definition into the main text to increase clarity.

We use the term teaching signal to mean exactly this definition throughout the manuscript. We suspect, a second source of confusion may come from ambiguity in regards to anatomical vs. functional definitions. We have attempted to emphasize that the definition of teaching signal is a functional one, not an anatomical one (as is the case for 'prediction' – predictions are functionally defined, not anatomically – hence it makes sense to ask questions of the form "what are the potential sources of predictions" etc.). A teaching signal is a signal that is compared against a prediction (the 'ground truth' the prediction is compared and trained

against). In different circuit implementations of predictive processing, different inputs function as predictions and teaching signals. In the hierarchical implementation, activity of the internal representation neuron at the lower level serves as a teaching signal for a prediction signal that is formed by the internal representation neuron from the higher level. In the non-hierarchical implementation, external inputs from the thalamus or other cortical areas serve as teaching signals for the respective prediction signals that are formed by internal representation neurons. This terminology is most intuitive when thinking from the perspective of a prediction error neuron, since a prediction error neuron computes the difference between two inputs signals – one of which functions as a teaching input, and the other one as a respective prediction. Hence, lines 428-430 specify that layer 5 input onto prediction error neurons of layer 2/3 serves as a teaching input.

> ***Reviewer #2 (Public review):***
>
> *Vasilevskaya and Keller set out to experimentally distinguish between two variants of predictive processing: a hierarchical and a non-hierarchical variant. The hierarchical variant assumes a hierarchical organization in which internal representation neurons (believed to be a subset of layer 5 excitatory neurons) serve as a source of a teaching signal for local prediction error neurons as well as for the next higher level of the hierarchy, while simultaneously providing prediction signals to the preceding lower level. In contrast, the non-hierarchical variant posits that these layer 5 internal representation neurons provide local predictions to layer 2/3 prediction error neurons.*
>
> *The interaction between internal representation neurons and prediction error neurons differs fundamentally between the two variants. In the hierarchical variant, internal representation neurons excite positive prediction error neurons and inhibit negative prediction error neurons, while at the same time being inhibited by positive prediction error neurons and excited by negative prediction error neurons. In the non-hierarchical variant, this pattern of connectivity is reversed.*
>
> *This work is very exciting, timely, and carefully executed. The authors functionally, and later molecularly, identify layer 2/3 prediction error neurons in V1 and probe their interactions with genetically defined neuron types in cortical layers 5 and 6 using optogenetics. They demonstrate that the functional influence of putative prediction error neurons in layer 2/3 onto layer 5 is incompatible with the hierarchical variant, whereas the influence of layer 5 onto putative prediction error neurons in layer 2/3 is incompatible with the non-hierarchical variant. They then test an alternative hypothesis, in which layer 2/3 responses resemble prediction errors with respect to perturbations of artificial layer 5 activity patterns. To investigate this, they designed an experiment in which optogenetic activation of L5 IT neurons was closed-loop coupled to the mouse's locomotion speed in the absence of visual feedback, allowing them to probe the causal influence of L5 activity on layer 2/3 responses.*
>
> *Finally, the authors hypothesize that their data are more consistent with a joint embedding predictive architecture (JEPA) and outline experimentally testable predictions arising from this framework.*

We thank the reviewer for their comments. We address them below.

> *While the work is overall convincing and significantly advances our understanding of the circuit-level implementation of predictive processing, there are a few weaknesses that should be addressed or discussed:*
>
> *(1) The authors define putative positive prediction error neurons as the 15% of neurons most responsive to grating onset and putative negative prediction error neurons as the 15% most responsive to visuomotor mismatch. While this selection would be expected to*

> *overlap with negative and positive prediction error neurons, the criterion is not sufficiently stringent (independent of the exact percentage chosen). In particular, classification of a neuron as a prediction error neuron should ideally be accompanied by evidence that it does not exhibit a significant increase in activity when the prediction matches the sensory input or teaching signal.*

We understand the reviewer's intuition. We can indeed use other stimuli to identify prediction error neurons, like the relative suppression of responses in closed-loop running onset vs open-loop running onsets. This was the reason behind including Figure S1, to show that our selection results in expected pattern of running onset responses. We don't typically use the running onset responses as they have an additional confound we have not fully understood. This is that running onset always tends to result in an increase of calcium activity in all neurons. This could have a variety of reasons: A) Contamination of hemodynamic occlusion signals (blood vessels tend to constrict at running onset, making it appear like an increase in calcium activity) – see Yogesh et al., 2025. B) The virtual coupling in our VR is not good enough to provide a true "closed loop" experience. Humans typically notice lags of larger than 30ms – in our VR it is approximately 100 ms. C). Running onset in head-fixed animals is not accompanied by a vestibular input. D) Predictive processing is wrong. We tend to think it is a combination of the three.

We can also use combinations of the two criteria to select neurons - if the reviewer has a specific selection criteria in mind (top XXX% MM responsive AND top XXX% closed-loop suppressed, etc.) we are happy to repeat the analysis for that specific set of criteria, but the fundamental problem that we are using a functional response to select these neurons does not go away. We know that our functional selection criteria mean we select a population of neurons that is enriched for prediction error neurons. If the enrichment is too weak, we would expect to find no effects in terms of functional influence. It is hard to explain, however, how a weak enrichment could result in a strong effect on functional influence. Our arguments in more lengthy form, for why the visuomotor mismatch is a good stimulus to identify negative prediction error neurons can be found here: Attinger et al., 2017; Jordan and Keller, 2020; Leinweber et al., 2017; O'Toole et al., 2023; Vasilevskaya et al., 2022; Zmarz and Keller, 2016.

> *(2) The authors "speculate that the prediction error responses in layer 2/3 may not be computed with respect to sensory input, but with respect to layer 5 activity as a teaching signal." However, it is unclear how this perspective differs from earlier statements in the manuscript. In the Introduction, the authors note that "these signals, typically referred to as sensory signals, we will refer to as teaching signals," and later describe the hierarchical variant as one "in which internal representation neurons act as a source of the teaching signal." Given this framing, it is difficult to identify what is conceptually novel in the updated view. Is the key distinction that layer 2/3 neurons are now proposed to generate predictions in an internal representation space rather than in sensory input space, as briefly suggested in the Discussion? Or are the authors introducing a distinction between an external (sensory) and an internal (cortical) teaching signal? If so, this distinction should be made explicit. Clarifying this point would considerably strengthen the manuscript.*

There might be a misunderstanding regarding our usage of the term teaching signal. In hierarchical predictive processing the teaching signal is typically referred to as a sensory signal, as e.g. in: "prediction error neurons compare predictions to sensory input". In non-hierarchical predictive processing, or far away from the sensory input (think prefrontal cortex), or for cross-modal interactions "sensory" input is misleading. Also, in non-predictive-processing type models (like JEPA), sensory input has a different functional role. Thus, we operationally define teaching signal as the signal that is compared against the prediction by prediction error neurons.

The two primary options we are comparing are:

(1) Is the teaching signal to the L2/3 comparator a bottom-up input to V1 (as one would expect in predictive processing).

(2) Is the teaching signal to the L2/3 comparator L5 input (as one would expect in JEPA).

Our data argue in favor of option 2. We have rephrased parts of the manuscript to try to make this clearer.

> *(3) The authors propose that "L2/3 neurons predict L5 activity, hence making predictions in the internal representation space rather than the input space," and further suggest that, since both deep and superficial cortical layers receive thalamic input, the cortex may function like a JEPA. This idea appears closely related to the model introduced by Nejad et al. (2025), which effectively implements a JEPA-like architecture: L5 activity serves as a target against which L2/3 predictions are compared in a selfsupervised manner, with both L5 and L2/3 (via L4) receiving thalamic input. It would be helpful for the authors to clarify how their framework differs from that model, and to specify the key conceptual or mechanistic distinctions between the present proposal and the approach described by Nejad et al.*

The two proposals indeed share similarities in assuming that bottom-up input for both L2/3 and L5 arrives from thalamus, and that representations formed in L2/3 are used for predicting the activity of L5. However, there are a few important differences between the JEPA implementation proposal formulated here and the Nejad et al. model.

(1) There is no proposed mapping of computations in the Nejad et al. model onto different JEPA networks. We assume that the suggested mapping would be L4 and L5 as encoder networks, and L2/3 as a predictor network? In that case, it is different to our proposal, in which L2/3 is part of the encoder network.

(2) Our proposal contains explicit prediction error neuron cell types within L2/3, while prediction errors in Nejad et al. are encoded in the gradients, and the layer origin of these signals is hypothesized to be L5 (‘the learning-driving error signal originates in L5’). Hence, also the role of L5-L2/3 connection is distinct between the two proposals. In Nejad et al. this connection serves the role of error propagation and update for predictions in L2/3, while in our proposal this connection contains teaching signal (target representations) that are compared to predictions within L2/3. Similarly, the functional role of L2/3-L5 connection is also different, since in Nejad et al, it is supposed to carry predictions of L5 activity, whereas in our proposal we expect it to drive plasticity in L5 encoder.

(3) The difference outlined above also makes it evident that the two proposals should differ in how deep and superficial layers are expected to influence the activity of one another. Indeed, the proposal in Nejad et al. is based on the cortical column idea, and according to eq. 2 and 3 in the Methods, activity in L5 is a function of activity in L2/3, while activity in L2/3 is not a function of activity in L5. Our proposal is based on idea of layers forming parallel networks, where horizontal communication is the dominant mode of cortico-cortical interactions, and activity in deep layers serve as a teaching signal for L2/3. In our case, we expect the opposite - that activity in L2/3 depends on activity of L5, while activity of L5 is not immediately dependent on activity of L2/3 (only via plasticity route). This led us to propose one of direct tests for our framework – silencing L2/3 in a familiar setting should result in no immediate changes to L5 activity and behavior of the animal.

(4) The proposal in Nejad et al. relies on input reconstruction or variance maximization within the L5 autoencoder network to avoid collapse. Instead, our proposal has no reconstruction objective.

(5) Lastly, there is a time delay between inputs to L5 and L2/3 that is proposed in Nejad et al., while this is not something inherent to our proposal.

We expect that the most useful future models should move beyond JEPA, with the emphasis on nonhierarchical models capable of operating on arbitrary graphs. We think cortex functions according to principles of a JEPA (predictions in latent space), and that the role of cell types and the exact computational organization remain to be constrained.

REFERENCES

Attinger, A., Wang, B., Keller, G.B., 2017. Visuomotor Coupling Shapes the Functional Development of Mouse Visual Cortex. Cell 169, 1291-1302.e14. *https://doi.org/10.1016/j.cell.2017.05.023*

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P.H., Buchatskaya, E., Doersch, C., Pires, B.A., Guo, Z.D., Azar, M.G., Piot, B., Kavukcuoglu, K., Munos, R., Valko, M., 2020. Bootstrap your own latent: A new approach to self-supervised Learning. *https://doi.org/10.48550/arXiv.2006.07733*

Jordan, R., 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*

Leinweber, M., Ward, D.R., Sobczak, J.M., Attinger, A., Keller, G.B., 2017. A Sensorimotor Circuit in Mouse Cortex for Visual Flow Predictions. Neuron 95, 1420-1432.e5. *https://doi.org/10.1016/j.neuron.2017.08.036*

Mohammadi, A.G., Halvagal, M.S., Zenke, F., 2025. Understanding cortical computation through the lens of joint-embedding predictive architectures. *https://doi.org/10.1101/2025.11.25.690220*

O'Toole, S.M., Oyibo, H.K., Keller, G.B., 2023. Molecularly targetable cell types in mouse visual cortex have distinguishable prediction error responses. Neuron 111, 2918-2928.e8. *https://doi.org/10.1016/j.neuron.2023.08.015*

Saravanan, V., Berman, G.J., Sober, S.J., 2020. Application of the hierarchical bootstrap to multi-level data in neuroscience. Neurons Behav. Data Anal. Theory 3, *https://nbdt.scholasticahq.com/article/13927-application-of-the-hierarchical-bootstrap-tomulti-level-data-in-neuroscience*.

Vasilevskaya, A., Widmer, F.C., Keller, G.B., Jordan, R., 2022. Locomotion-induced gain of visual responses cannot explain visuomotor mismatch responses in layer 2/3 of primary visual cortex. *https://doi.org/10.1101/2022.02.11.479795*

Yogesh, B., Heindorf, M., Jordan, R., Keller, G.B., 2025. Quantification of the effect of hemodynamic occlusion in two-photon imaging of mouse cortex. eLife 14, RP104914. *https://doi.org/10.7554/eLife.104914*

Zmarz, P., Keller, G.B., 2016. Mismatch Receptive Fields in Mouse Visual Cortex. Neuron 92, 766–772. *https://doi.org/10.1016/j.neuron.2016.09.057*

https://doi.org/10.7554/eLife.110827.2.sa0