Recurrent Neural Networks reveal distinct signal flow in visual cortex in response to novel stimuli

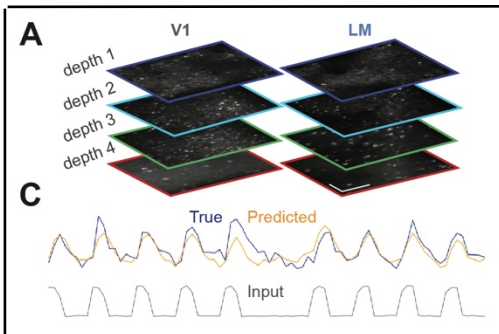
Predictive coding is a theory of brain function in which the brain constantly generates predictions about the environment and updates the predictions if they deviate from the actual sensory inputs[1]. Theoretical models of predictive coding have relied on communication across brain areas for generating and updating predictions[2]. Therefore, to investigate the neural circuit mechanisms that underlie predictive coding, we need to study signal flow across cortical areas while animals make and update predictions. Here we addressed this question by using Recurrent Neural Networks (RNN's) to study cortical layer and area interactions during a visual task which involved expectation violations. In brief, Multiplane Mesoscopes[3] were used to simultaneously image 4 cortical depths of 2 visual areas (V1 and LM), while mice received a stream of visual stimuli. Three mouse lines, tagged for excitatory and 2 inhibitory subpopulations, VIP and SST, were imaged. Each session included regular image presentations that were occasionally omitted ("omissions"). The same image set was presented for several days (familiar sessions), before switching to a novel image set (novel sessions). We have previously demonstrated that in familiar (but not novel) sessions, image anticipatory responses as well as omission-evoked responses evolve in VIP neurons, suggesting that certain neurons in visual cortex may represent expectation violation signals[3-6]. In the current study, we leveraged RNNs to investigate the modification of functional connectivity across neural populations[7]. In particular, we studied how functional connectivity across the depths of V1 and LM changes in novel sessions. Our findings demonstrate that RNN input and recurrent weights are modified in novel sessions, distinctly across cell types. These results give insight into how bottom-up inputs and inter-area connections are modified in excitatory and inhibitory subpopulations when animals receive unexpected novel stimuli, and hence, illuminate our understanding of the neural pathways involved in predictive coding.

RNN training. We trained an RNN on the neural recordings from each session. The input to the model was a sequence of 9 image presentations and 1 omission. The output of the model was the calcium trace of recorded neural activity across 8 cortical planes (Fig. 1). We divided data from each session into 10 unique batches, and trained 1 RNN per batch; hence, each session yielded 10 RNN's with 10 unique sets of weights. Having multiple RNN's serves two purposes. First, it allows potential changes in connectivity dynamics throughout the experiment to be recorded in different RNN's models. Second, it adds more weights into the weight distribution, improving the statistical rigor. This methodology is then run across multiple random initialization seeds and multiple recorded experiments.

Results. To gain insight into how functional connectivity changes as a result of novel stimuli, we compared the weight magnitude of RNN units between familiar and novel sessions. RNN input weights can be interpreted as the strength of the bottom-up pathways that carry stimulus information; while the recurrent weights inform us about the strength of connections across cortical areas/layers. Our results indicate that following novel images, the overall connectivity strength increases in the excitatory and VIP populations; while it decreases for the SST population (Fig. 2B,C). These findings are compatible with how population averages are modified for each cell type in response to novelty (Fig. 2A). Novel images differentially modified connectivity strength across cortical planes, particularly in excitatory and VIP populations. In the excitatory population, input and recurrent weights were primarily increased in LM, deep layer 4 (Fig. 2B,C; top). Interestingly, recurrent weights in V1 were largely unaffected by novel stimuli (Fig. 2C; top; within V1). In the VIP population, the most noticeable effect was that feedback connectivity (LM to V1) strengthened in superficial layers 2/3, yet weakened in layer 4 (Fig. 2C; bottom;

LM to V1). The reconnect connectivity of SST population was broadly modified across cortical planes (Fig. 2C; middle); however, the input weights were preferentially increased in layer 4 of V1 and LM (Fig. 2B; middle).

Discussion. Using RNN's as a proxy model of the mice visual system, we demonstrate distinct changes in the functional connectivity of excitatory and inhibitory neurons across cortical layers and areas following exposure to novel stimuli. Our results suggest that as a result of novelty, connections among excitatory neurons increase predominantly in deep layer 4 of LM, a higher order visual area; while connections among SST inhibitory neurons decrease broadly across V1 and LM layers. VIP neurons showed a distinct effect between feedforward and feedback connections: while feedforward connections increased rather broadly across layers, feedback connections strengthened only onto superficial layers of V1, and weakened onto layer 4. The overall change of weights in excitatory, SST and VIP populations confirms the classical disinhibitory VIP-SST-Excitatory circuit. Previous evidence has suggested that feedback projections from higher order cortical areas onto V1 L2/3 carry prediction signals[8,9]. Our findings are in line with the previous evidence and provide further insight into the cortical circuitries that underlie the representation and updating of prediction signals. Future investigation from a dynamical system standpoint of the RNN will allow us to understand possible mechanisms that are used as learning rules to update internal representations.



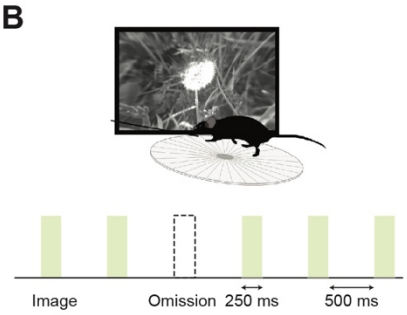


Figure 1. A) Multiplane Mesoscopes were used to image 4 cortical depths in V1 and LM. **B)** Mice were presented with a sequence of images and occasional omissions. **C)** RNN's were trained to predict the activity of individual neurons from visual inputs.

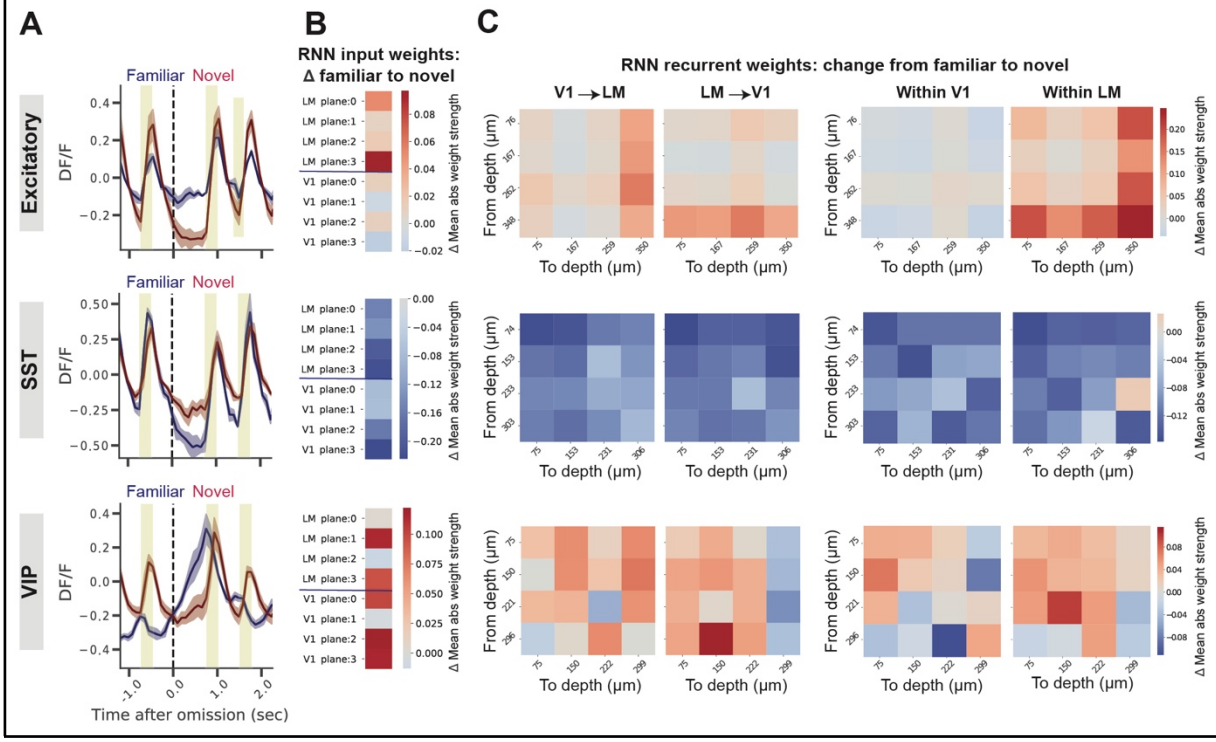


Figure 2. A) Population averages of individual cell types in response to familiar (blue) and novel (red); yellow boxes: Image presentations. **B-C)** RNN input weights and recurrent weights of neurons in different cortical depths. Values show the average of the absolute magnitude of the weights; hence reflect both the center and variance of the weight distribution for each plane.

1) Keller & Mrsic-Flogel, Neuron, 2018. **2)** Rao & Ballard, Nature Neuroscience, 1999. **3)** Orlova*, tsyboulski*, Najafi*, et al, bioRxiv, 2020. **4)** Garrett et al, eLife, 2020. **5)** Jordan & Keller, Neuron, 2020. **6)** Hamm & Yuste, Cell Reports, 2016. **7)** Andalman et al, Cell, 2019. **8)** Marques et al. Nature Neuroscience, 2018. **9)** Keller et al, Nature, 2020