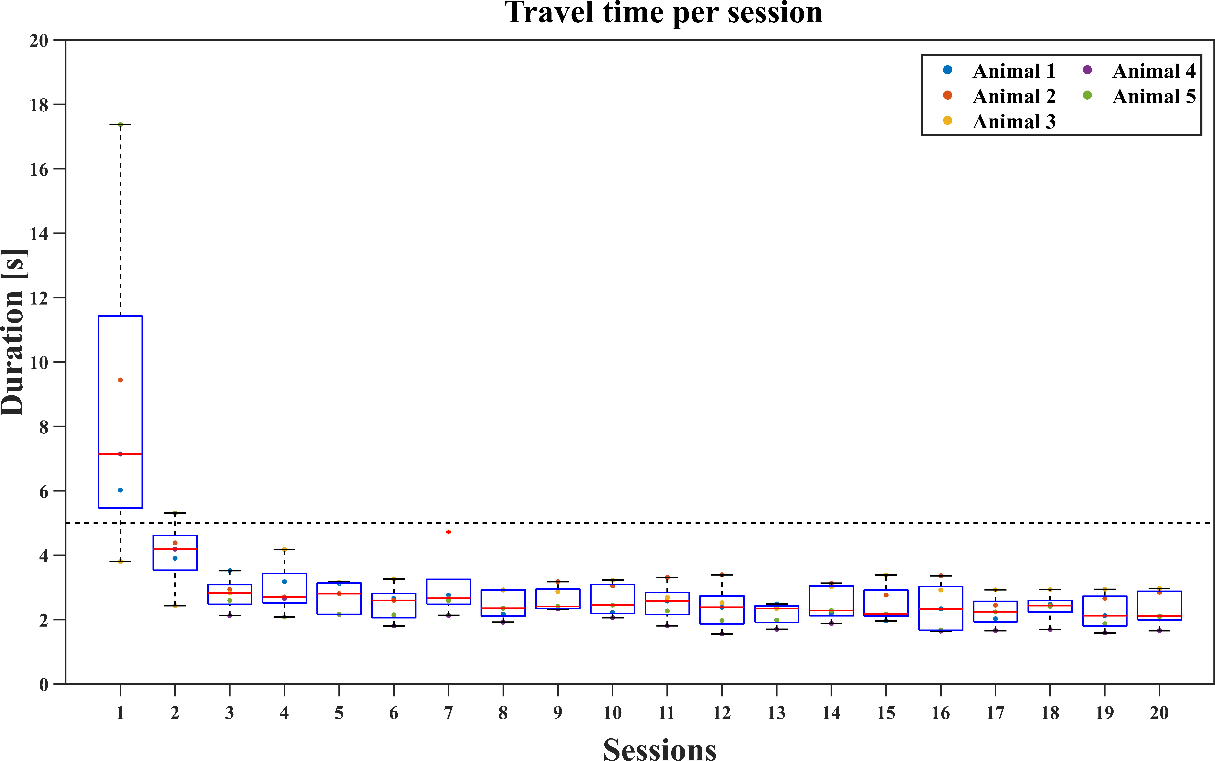
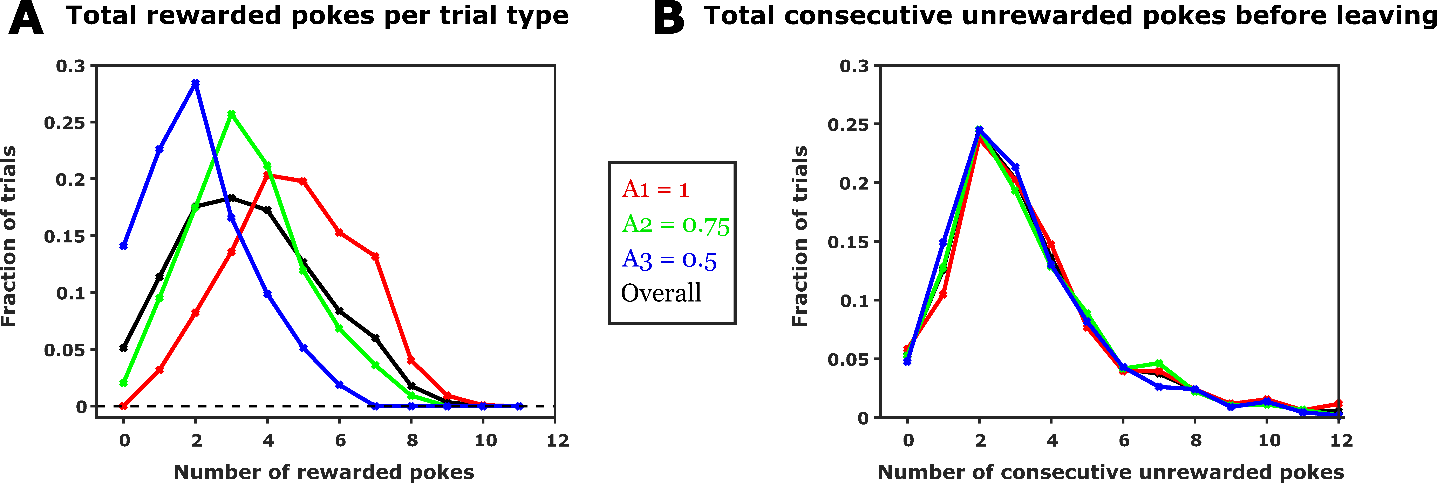


**Figure 1: Sample histology image of the frontal region A (FrA).** The histology slice is taken 4.85 mm anterior to Bregma. The electrode location from the staining can be seen 1.5 mm lateral with a cortical depth of 1.29 mm approximately.



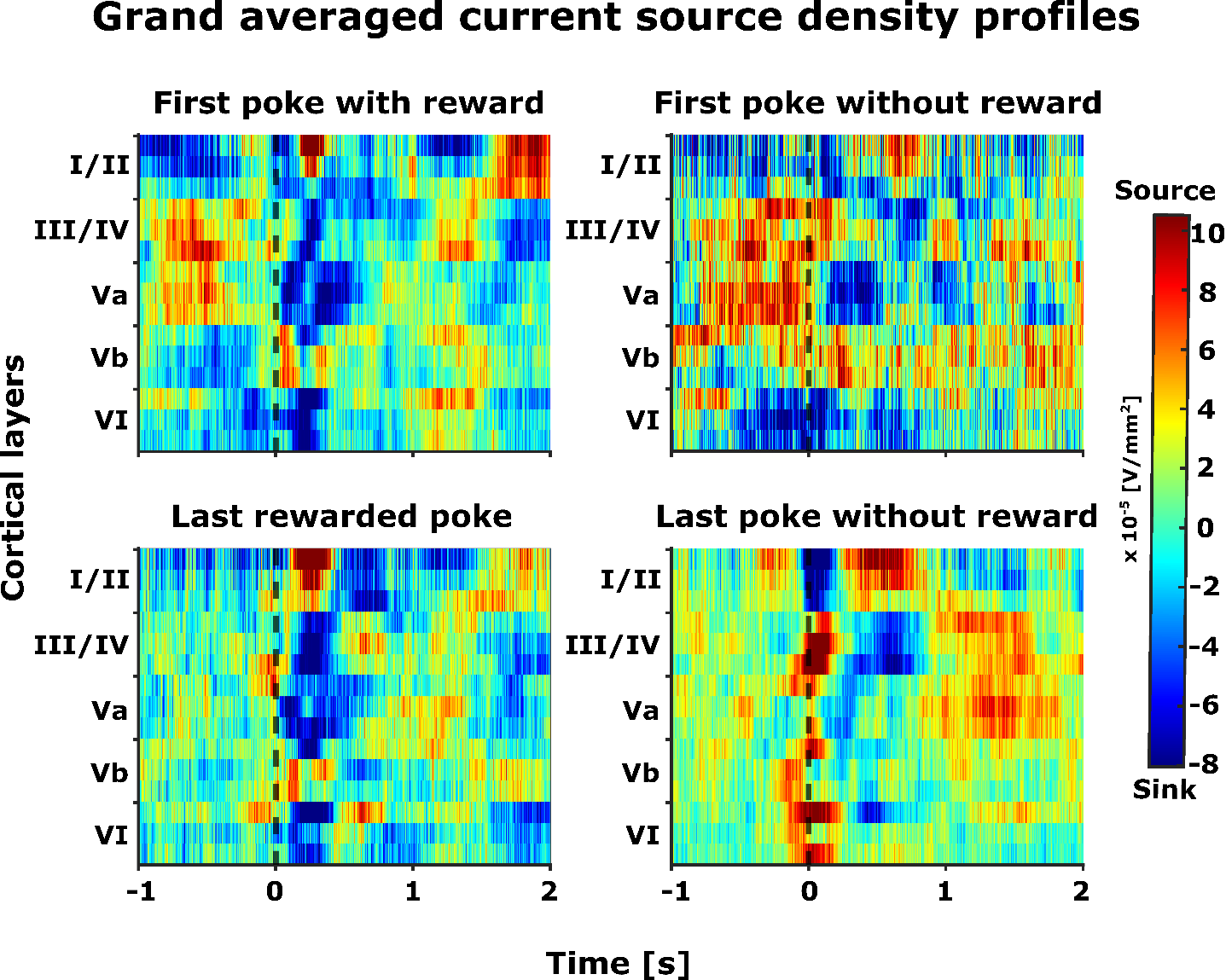


**Figure 2: Performance of the foraging behaviour. A** shows the median ± SD of travel time per session remains consistent and less than 5 seconds. Travel time is calculated as the time between the end of last poke in a trial to the start of the first poke in the succeeding trial. **B** shows the total number of rewarded pokes across trials for different starting probabilities (A = 1, 0.75 and 0.5). **C** shows the total consecutive unrewarded pokes before leaving a spout in each trial having different starting probabilities.

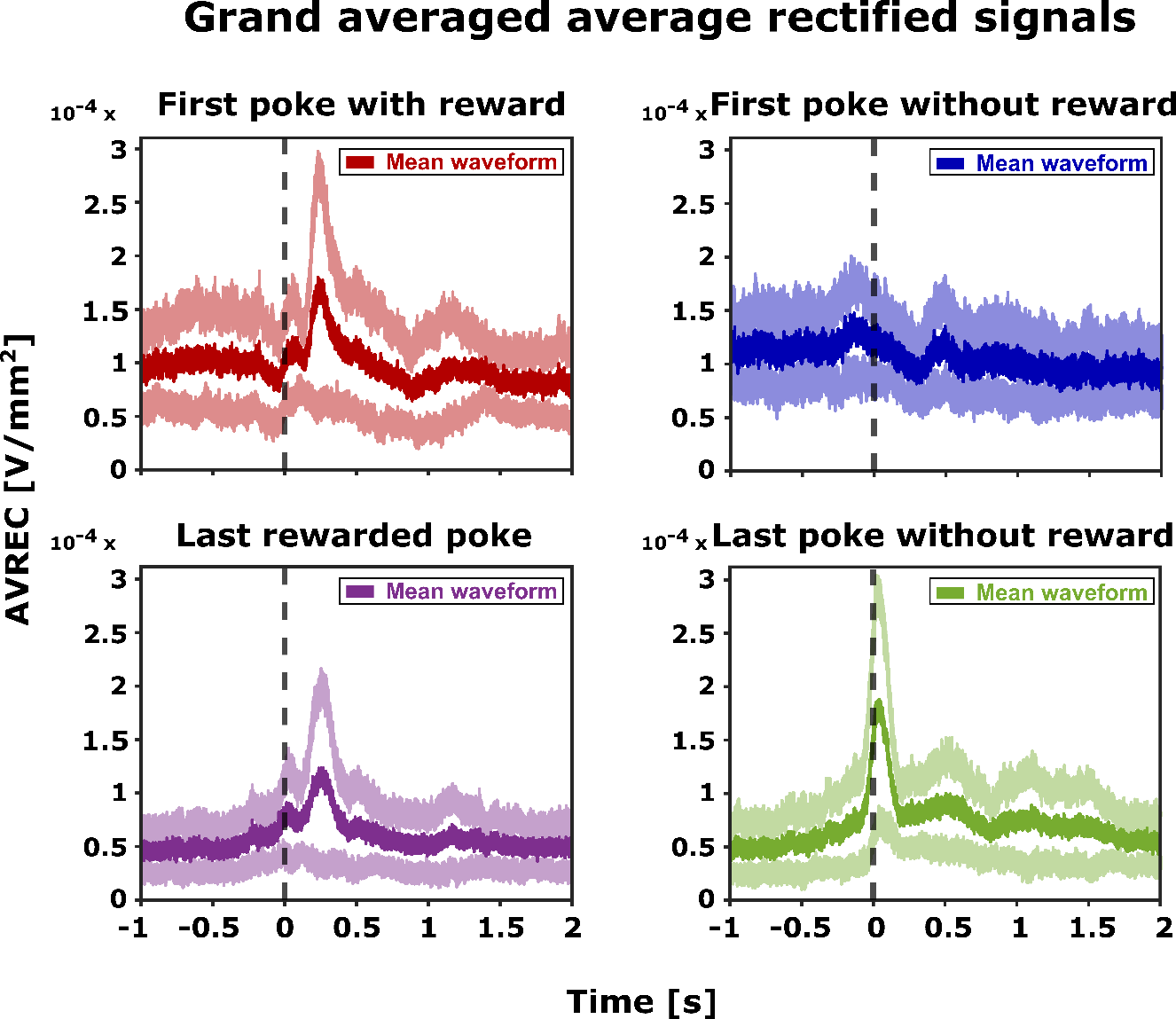
Points to discuss:

1. How travel time is used as an indirect measure to interpret animal’s learning? The rationale is that lesser and consistent travel time indicate that the animal is goal-directed and not randomly exploring. Based on Fig.2A, the consistent and lower than 5s travel time indicates that after each trial, they were directed towards the other spout and not randomly exploring the arena. Hence, sessions 5-20 were considered for further behavioural and electrophysiological analyses.
2. Animals make inference-based decisions.

Following the experimental design (Methodology figure showing the exponential decrease of rewards for different starting probabilities), Fig.2B shows that the distribution of total rewards shifts to right for trials starting with higher initial reward probability. This indicates that the if the animal essentially makes same number of pokes in each trial, it will receive higher number of rewards for trials starting with higher initial reward probability. However, Fig.2B shows that the animals irrespective of the starting reward probability and the total number of rewards received in a trial maintains a consistent number of consecutive unrewarded pokes before leaving a particular spout. This indicates that the animals infer the hidden probabilities based on their learning of the task (explain stimulus bound and inference bound decision making in discussion).



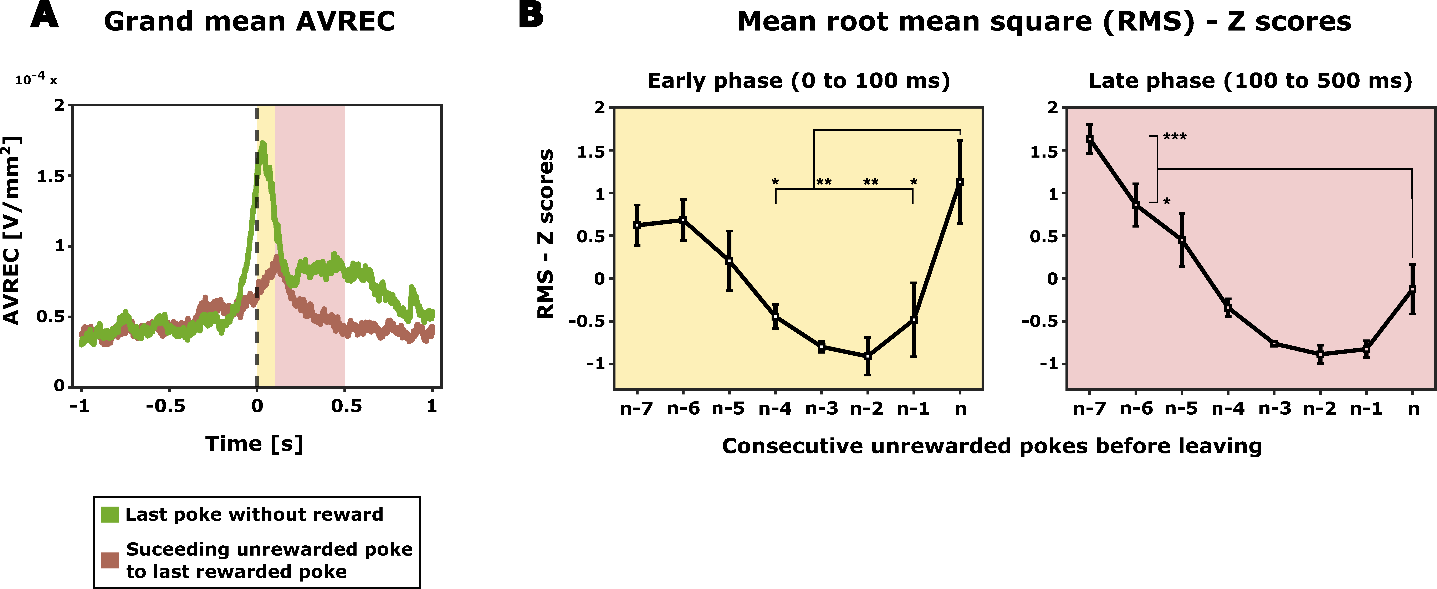
**Figure 3: Grand averaged current source density (CSD) profiles (n=5)** – The selected epochsrepresent -1 to +2 seconds from the end of the poke (black dashed line, t=0). The selected time interval was taken for four different events (pokes) and its corresponding consequence (reward): (top left) first poke with reward, (top right) first poke without reward, (bottom left) last rewarded poke, and (bottom left) last poke without reward.



**Figure 4: Grand averaged AVREC (n=5)** – Mean average rectified waveform (dark) along with its standard error (light) is plotted for selected time intervals (epochs). The selected epochsrepresent -1 to +2 seconds from the end of the poke (t=0). AVREC were taken for four different events (pokes) and its corresponding consequence (reward): (top left) first poke with reward, (top right) first poke without reward, (bottom left) last rewarded poke, and (bottom left) last poke without reward.

Points to discuss:

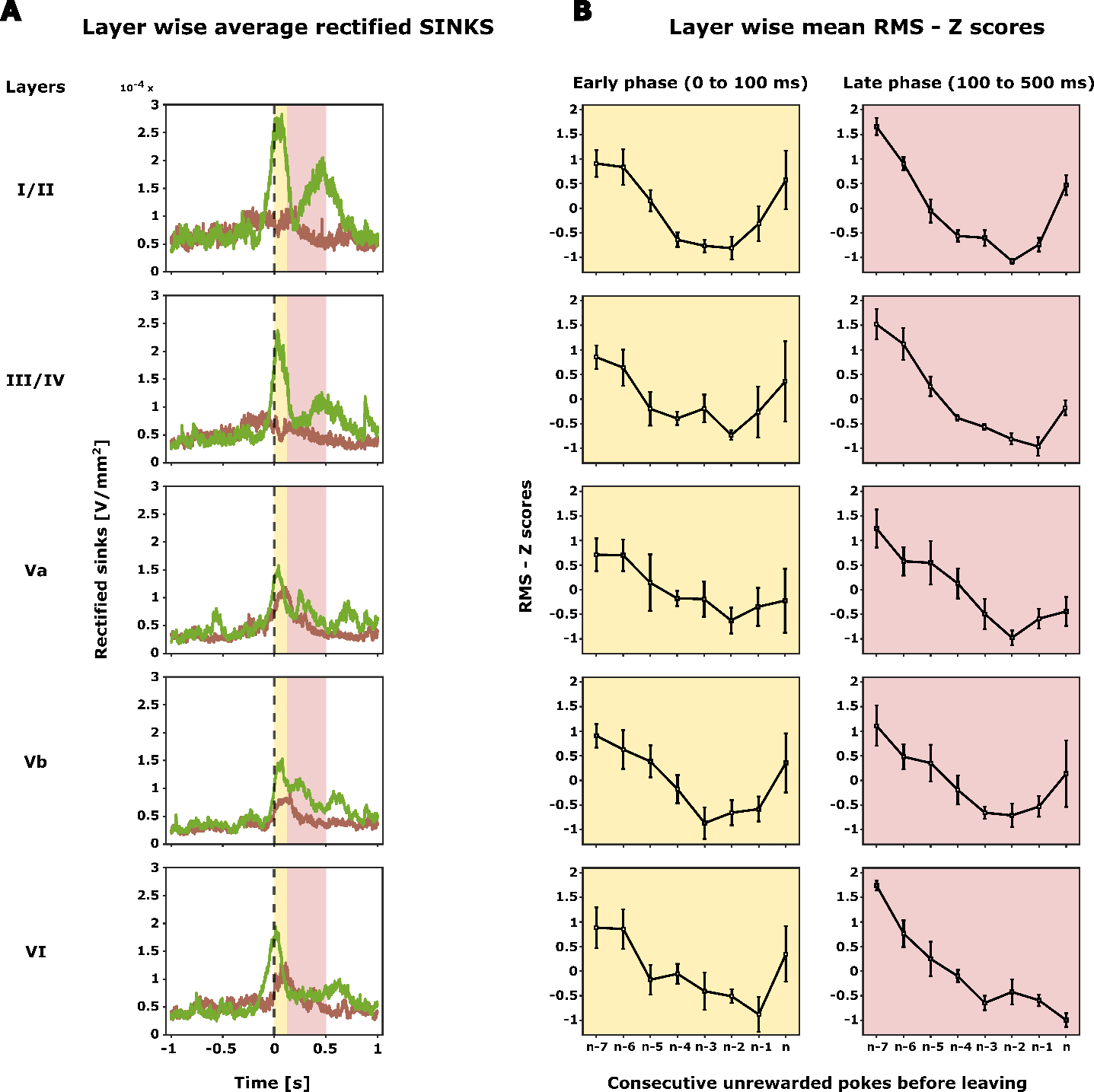
1. Distinct motor and reward related activity patterns. Distinct neural representations within frontal field A encode both the poke and the ensuing reward. The distinct activity patterns for different epochs seen in Fig.3 i.e., first rewarded poke, first unrewarded poke, last rewarded poke and last poke show that the frontal field A encodes not just the motor activity but also the reward related activity (expectation and prediction error).
2. The average rectified signals (AVREC) from the CSD profiles also show distinct motor and reward related activity patterns (Fig.4). Further, it can be seen that in frontal field A, the encoding of the expected (towards the end of pokes) and received reward (subsequently) shows different activity patterns during the exploitation (first poke until last rewarded poke) and exploration (last poke) phases of the trial.



**Figure 5: Shift from exploitation to exploration**. **A** – shows the distinct activation pattern of grand AVREC of first unrewarded poke after last rewarded poke (brown) and the last poke without reward (green). Based on the grand AVREC data (A), two distinct time intervals (epochs) were chosen for RMS computation: early phase (0 – 100 ms, yellow), and late phase (100 – 500 ms, light pink). **B** – The average root mean square (RMS) Z score for unrewarded pokes between last rewarded poke and the last poke (n) before disengaging from the current spout (in this figure, we show a scenario of 7 consecutive unrewarded pokes where n-7 is the first unrewarded poke after last rewarded poke while n represents the last poke (unrewarded)). How to write the statistics part?

Points to discuss:

1. What makes the last poke the last poke? Now that we have confirmed that the frontal field A encodes distinct activity patterns for different poke phases (Fig.3 and Fig.4), it is essential to understand how the animal decides the last poke in a trial. This is important because, in a trial from a reward stand point, the animals may encounter multiple consecutive unrewarded pokes after the last rewarded poke. Hence, the last poke also being an unrewarded poke needs to be distinguished from its preceding unrewarded pokes.
2. Fig.5A evidently shows that the last unrewarded poke has a distinct activity pattern compared to that of first unrewarded poke succeeding the last rewarded poke. Further, during the exploitation phase (first unrewarded poke after last rewarded poke), extended persistent activity patterns are observed after 100 ms following the nose poke (reward evaluation). Conversely, at the onset of the exploration phase (last poke), heightened frontal activity is evident during the animal's decision-making, occurring within 100 ms after the nose poke.
3. Now, based on the grand AVREC (Fig.5A), two different phases were selected i.e., early phase (0-100 ms from the end of the poke, Fig.5A yellow phase) and late phase (100-500 ms from the end of the poke, Fig.5A light pink phase). In order to see the evolution of change in activity patterns towards the last poke, we computed the z-scores of root mean square (RMS) of AVREC from each of these phases for all the unrewarded pokes between last rewarded poke and last poke (Fig.5B and 5C). Over the transition from exploitation to exploration (n-7 to nth poke), there is an initial phase where the overall frontal activity decreased (n-6 to n-2) and then increased just before the animal decides to leave the spout (n-2 to nth poke).



**Figure 6: Layer specific frontal motor and reward related activity**. **A** – The grand averaged rectified sinks (n=5) was computed for all the identified layers from the grand CSD profile (Fig.3). The selected epochsrepresent -1 to +2 seconds from the end of the poke (t = 0). The selected time interval was taken for first unrewarded poke after last rewarded poke (brown) and last poke without reward (green). Based on the averaged sinks (A), two distinct time intervals (epochs) were chosen for RMS computation: early phase (0 – 100 ms, yellow), and late phase (100 – 500 ms, light pink). **B** – The layer wise average root mean square (RMS) Z score was calculated from the average rectified sinks for unrewarded pokes between last rewarded poke and the last poke (n) before disengaging from the current spout. (Similar to Fig.5B, in this figure, we show a scenario of 7 consecutive unrewarded pokes where n-7 is the first unrewarded poke after last rewarded poke while n represents the last poke (unrewarded)).

Points to discuss:

1. Comparison of layer-wise averaged sink activity indicates a transition of activity from deeper to upper cortical layers once the reward expectation at the current spout decreased (add last rewarded poke). Shortly before the decision of the animal to explore the other spout, particularly layers I/II and III/IV showed two prominent activity peaks directly at the retraction of the spout (0-100 ms) and afterwards (100-500 ms) (Fig.6A).
2. During the phase of switching from exploitation to exploration, activity decreases in all cortical layers. At the beginning of the exploration phase, the upper layers I/II and III/IV, as well as Vb, are active. The increase is mainly observed in the early phase following the nose poke (evaluation), and then with regard to the evaluation of the absence of rewards (Fig.6B).