# Discussion

## Gerbils perform inference bound decision making

## Performance of the foraging behaviour

In order to see if the animal has learnt the foraging task and reduce the possibility of random exploration, we looked at their travel time. Here, travel time is defined as the time between the end of last poke in a trial and the start of the first poke in the succeeding trial (Fig1B, Eq.2). Lesser and consistent travel time indicate that the animal is goal-directed and not randomly exploring. The consistent and lower travel time indicates that after each trial, the animals were directed towards the other spout and not randomly exploring the cage (Fig.2A). Based on this, only sessions from 6 to 20 were considered for further behavioural and electrophysiological analyses.

### Animals make inference-based decisions

To evaluate how well the animal learnt to make decisions during the exploitation-exploration dilemma, we focussed on the time point where the animal decided to shift from exploitation to exploration. This is because every trial begins with exploiting the current spout for rewards and slowly as the rewards gets exhausted, the important decision is to decide when to leave the current spout to explore the other. This can be interpreted by the consecutive number of unrewarded pokes the animal makes before leaving the current spout.

Following the experimental design (Fig.1C), Fig.2B shows that the distribution of total rewards shifts towards right for trials starting with higher initial reward probability. This indicates that the if the animal makes same number of pokes in each trial, it will receive a greater number of rewards for trials starting with higher initial reward probability. On the contrary, this distinction between trials starting with different starting reward probabilities gets lost when it comes to the consecutive unrewarded pokes made before switching (Fig.2C) In other words, it 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 consistency in decision making may reflect the possibility that the animals form an inference about the hidden reward structure based on their learning of the task.

## Distinct motor and reward related activity patterns in the frontal field A

To confirm if the target region FrA encodes distinct activity patterns that reflects the animal’s behaviour, four different time points that are unique and represents crucial stages during the foraging session were selected based on the animal’s performance (Fig1B and Fig2). These are first poke (rewarded and unrewarded), last rewarded poke and the last poke. The time between first poke until last rewarded poke can be considered as the exploitation phase because during this phase, even after experiencing unrewarded pokes, the animal insists on staying on the same side expecting more rewards. The last rewarded poke is a critical point as it represents the end of exploitation phase and beginning of transition towards exploration. This is because after the last rewarded poke, the animal slowly starts to alter its expectation and move towards exploration phase which is best represented by the last poke in a trial. Fig.3 shows distinct spatiotemporal neural activity within frontal field A that encode both the poke (motor) and the ensuing reward related information. The selected epochs represent -1 to +2 seconds from the end of the poke (black dashed line, t=0) to compare the neural activity during decision-making phase. Further, the differential reward related activity patterns helped identify and distinguish infragranular layers from superficial layers in the laminar recordings and perform channel-layer specification.

In order to visualize the overall frontal cortical activity, the CSD signals were rectified and averaged across the laminar electrodes to lose the spatial information. The average rectified signals (AVREC) from the CSD profiles also show distinct motor (Fig.4, peak around end of the poke (t=0)) and reward related signals (Fig.4, peak around 250 ms from the end of the poke). From the perspective of expectation and evaluation of reward, it can be seen that in FrA, 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. Both the rewarded pokes (first rewarded and last rewarded pokes) show a similar waveform where there is an early peak in amplitude immediately after the end of the poke that may encode for the expectation of a reward followed by a slightly increased peak occurring in less than 250 ms from the end of the pokes that may encode for the reward evaluation. On the other hand, in the unrewarded pokes (first unrewarded and last poke), the reward evaluation shows a dip followed by a later peak in amplitude around 500 ms that may encode for a prediction error from the deeper layers. Compared to all other pokes, the last poke which is the starting point of exploration phase where the animal decides to switch to the other spout, shows a heightened frontal activity immediately after the end of the poke (<100 ms). These distinct activation patterns for different time points seen in Fig.3 and Fig. show that the frontal field A encodes not just the motor activity but also the reward related activity (expectation and prediction error).

## Shifts in frontal activity patterns: Exploitation to exploration

### What makes the last poke the last one?

As we have confirmed that the frontal field A encodes distinct activity patterns for different phases (Fig.3 and Fig.4), it is essential to understand how the animal decides the last poke in a trial. This is essential as the animals may encounter multiple consecutive unrewarded pokes after the last rewarded poke (Fig.5A). Hence, the last poke also being an unrewarded poke needs to be distinguished from its preceding unrewarded pokes.

Fig.5B evidently shows that the last unrewarded poke has a distinct activity pattern compared to that of first unrewarded poke succeeding the last rewarded poke. In order to decode when the decision making exactly occurs, first 500 ms from the end of the poke was considered and two distinct phases were selected. Early phase represents the first 100 ms (Fig.5B, yellow region) and the late phase represents 100-500 ms (Fig.5B, light pink region) from the end of the poke. This way, we can verify if the decision making is happening immediately after the nose poke action or it also requires a reward evaluation information. When the animal is still at the end of exploitation phase (first unrewarded poke after last rewarded poke), less activity is observed at the early phase but extended persistent activity patterns are observed after 100 ms following the nose poke (reward evaluation) indicating that the animal is still focussed on getting a reward. Conversely, at the onset of the exploration phase (last poke), heightened frontal activity is evident within 100 ms after the nose poke (early phase) indicating a focus shift from the reward expectation towards the nose poking action. This shift in activity pattern makes the last poke unique and hints that the early phase could be the crucial phase where the primary decision to explore may get encoded.

### The evolution of frontal activity from exploitation to exploration

To get a deeper understanding of the evolution of change in frontal activity patterns from exploitation to exploration, we focussed on the individual pokes. We computed and compared the root mean square (RMS) of the AVREC signal for all the unrewarded pokes between last rewarded poke and last poke (Fig.5B and 5C) and z-normalized it within each animal. Over the transition from exploitation to exploration (Fig.5B, n-7 to nth poke), during the early phase, the overall frontal activity initially decreased (n-6 to n-2) and then increased just before the animal decides to leave the spout (n-2 to nth poke) (One way ANOVA with Bonferroni correction, \* p < 0.05, \*\* p < 0.01, \*\* p<0.001). However, during the late phase, as it mainly encodes for reward evaluation, it constantly decreases as all the pokes considered here are unrewarded. The non-linear trend of activity change during the early phase indicates the involvement of a complex processing strategy underlying the inference-based decision-making behaviour seen in the animals (Fig.2C).

## Layer specific motor and reward related activity patterns in the frontal field

Five distinct cortical layers were identified from the current source profiles of the laminar recordings (Fig.3). The source signal was removed and only the sinks were considered to ensure that the signal is contributed only by layer specific local excitatory synaptic populations. Comparison of layer-wise averaged sink activity between the last poke and first unrewarded poke succeeding the last rewarded poke indicates that the large difference in activity patterns observed in overall frontal activity (Fig,5A) is mainly contributed by the large difference seen in the superficial layers (layer I/II and III/IV) (Fig.6A). Further, shortly before the decision of the animal to explore the other spout, particularly layers I/II and III/IV showed two prominent activity peaks in amplitude directly at the retraction of the spout (early phase, 0-100 ms) and afterwards (late phase, 100-500 ms). In the individual poke RMS analysis, similar to overall frontal activity, there is an initial decrease in all cortical layers. However, at the beginning of the exploration phase (towards the last poke (n)), the activity of upper layers I/II and III/IV, as well as Vb, increases. The increase is mainly observed in the early phase following the nose poke, and then with regard to the evaluation of the absence of rewards (Fig.6B). In the late phase, with the exception of layer I/II and layer III/IV which showed a steep increase in activity just before the last poke, all the cortical layers showed a decrease in activity across all the pokes similar to that of overall frontal activity.