# The flow of reward information through neuronal ensembles in the accumbens

Benjamin Arroyo[1], Enrique Hernandez Lemus[2,3], and Ranier Gutierrez*[1,4]

[1]Laboratory Neurobiology of Appetite; Department of Pharmacology, CINVESTAV, 07360, Mexico City, Mexico.

[2]Computational Genomics Division, National Institute of Genomic Medicine (INMEGEN), 14610, Mexico City, Mexico.

[3]Center for Complexity Sciences, Universidad Nacional Autónoma de México (UNAM), 04510, Mexico City, Mexico.

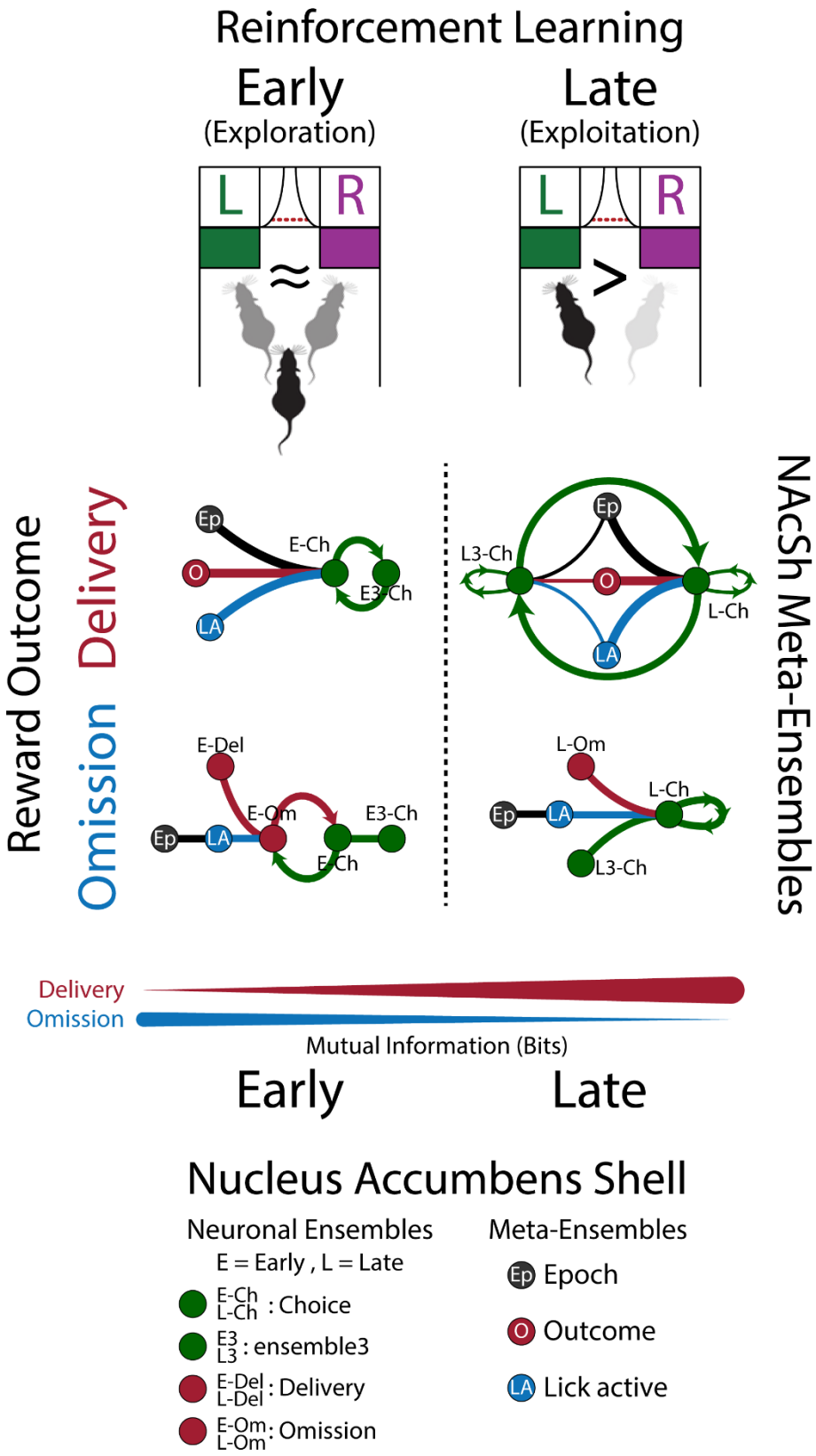
[4]Laboratory Neurobiology of Appetite; Centro de Investigación sobre el Envejecimiento (CIE), Cinvestav sede sur, México

*Correspondence should be addressed to R.G.

Email: ranier@cinvestav.mx

Number of pages: 78; Number of Figures: 8; Number of Tables: 0

**Graphical Abstract**



## Highlights

- Neuronal ensembles in the NAcSh act as dynamic modules to process different aspects of reward-guided behavior
- NAcSh ensembles dynamically change composition and interactions throughout reinforcement learning
- The NAcSh forms a highly connected network with neuronal hubs that facilitate an efficient reward information flow
- Reward triggers stronger interactions between ensembles and unifies network activity, while omission leads to less synchronization
- The NAcSh uses neuronal ensembles to process reward information and self-guide decision-making dynamically

**Summary**

The flow of reward information through neuronal ensembles in the nucleus accumbens shell (NAcSh) and its influence on decision-making remains poorly understood. We investigated these questions by training rats in a self-guided probabilistic choice task while recording single-unit activity in the NAcSh. We found that rats dynamically adapted their choices based on an internal representation of reward likelihood. NAcSh neurons encoded multiple task variables, including choices, outcomes (reward/no reward), and licking behavior. These neurons also exhibited sequential activity patterns resembling waves that peaked and dissipated with outcome delivery, potentially reflecting a global brain wave passing through the NAcSh. Further analysis revealed distinct neuronal ensembles processing specific aspects of reward-guided behavior, organized into four functionally specialized meta-ensembles. A Markov random fields graphical model revealed that NAcSh neurons form a small-world network with a heavy-tailed distribution, where most neurons have few functional connections and rare hubs are highly connected. This network architecture allows for efficient and robust information transmission. Neuronal ensembles exhibited dynamic interactions that reorganize depending on reward outcomes. Reinforcement learning within the session led to neuronal ensemble merging and increased network synchronization during reward delivery compared to omission. These findings offer a novel perspective of the flow of pleasure throughout neuronal ensembles in the NAcSh that dynamically changes its composition, with neurons dropping in and out, as the rat learns to obtain (energy) rewards in a changing environment and supports the idea that NAcSh ensembles encode the outcome of actions to guide decision making.

**Keywords**: Sucrose, feeding, reward, neuronal ensembles, foraging

## Introduction

The nucleus accumbens shell (NAcSh) is an integral part of the brain's reward system, a crucial hub that funnels the flow of pleasure and then broadcasts it throughout subcortical regions of the brain.[1,2] The NAcSh also plays a critical role in learning, motivation, and decision-making.[1,3,4] For example, NAcSh individual neurons encode incentive salience and motivation for rewards[5] and detect reward outcomes to guide decision-making.[4] However, the mechanism by which neurons in the NAcSh represent reward outcomes and guide choices is not fully understood. In parallel, the NAcSh is also considered a "sensory sentinel" for hedonic feeding,[6] which means that it is activated by the expectation of rewards but inhibited during sucrose consumption. Thus, consummatory behavior consistently inhibits NAcSh activity,[7–13] which gates feeding behavior.[14–16] It has been shown that $D1^+$ and $D2^+$ medium spiny neurons are inhibited during food consumption.[12,17,18] Likewise, feeding decreased glutamatergic tone in the NAcSh,[19,20] whereas optogenetic activation of glutamatergic inputs stops feeding and is rewarding.[16,21] In addition, a small group of NAcSh neurons tracks the hedonic value of sucrose and licking behavior (lick-coherent neurons).[11,18,22–24] However, how the NAcSh disentangles rewarding (hedonic) signals from the oromotor components of licking is also poorly understood.

We recently found that the activity of the NAcSh is dynamic, it changes depending on the behavioral task. For example, in a simple freely licking task, neurons in the NAc are organized into two major neuronal ensembles: One large neuronal ensemble exhibited a sustained inhibition, and a smaller ensemble was activated during sucrose consumption.[12,22] However, in a more complex task, in which animals have short access to tastants, such as a brief access test, the neurons in the NAc are organized in four functional neuronal ensembles that monitor the most relevant stimuli in the task [22] also see [18]. This suggests that the NAc's ensembles are the building blocks of neural circuits [25–27] that increase in number with task's complexity and play a dynamic role by helping animals track key task events to guide decision-making. We hypothesize that in a more complex task, such as a self-guided probabilistic choice task, the neuronal ensembles in NAcSh may reorganize even further. In this task, animals need to learn the probability of reward delivery for each lever and choose the most likely to deliver a reward.[28] This requires the animals to track their earlier choices, the outcome of those choices, and the current reward state of the environment. Notably, the rat receives no external cue to signal which actions lead to more reward. Instead, they must learn this on their own beliefs by building an internal

representation of the reward probabilities of each lever. Thus, this task is perhaps more challenging than one in which the rat is given external cues,[28,29] as it requires them to be more engaged in reinforcement learning and more vigilant in decision-making.

In this study, we performed multichannel recordings of the NAcSh activity while rats solved a self-guided probabilistic choice task with no external sensory cues. This task is a reinforcement learning paradigm in which rats learn to choose between two levers, each associated with a different probability of delivering a reward (3 drops of sucrose).[30] The outcome probability associated with each lever changed in blocks of 50 trials each, so the rat must adapt and constantly update their internal representation of reward probabilities. These features make this task ideal for disentangling the reward representation from the oromotor features evoked by licking and for a better understanding of how reward outcomes guide decision-making during a reinforcement learning task.

# Results

## Behavioral results

### Rats dynamically adjust their choices in response to reward probability and reward history

Seventeen rats were trained in a self-guided probabilistic choice task that closely resembled the reinforcement learning paradigm of the two-armed bandit task.[28,30] The task structure is illustrated in **Figure 1A**: Each trial started when the house lights turned on, signaling the rat to visit the center port and break the photobeam with its nose. After a 250-ms delay, a sound (Go signal) played, and both levers extended into the box. The rat could press either lever, with the sound not indicating which lever was more likely to yield a reward. The reward probabilities associated with the left and right levers were varied across blocks (Combination assayed: 10%, 50%, or 90%; **Figure 1B**). These probabilities changed every 50 trials with no explicit cue or sensory feedback. After pressing the lever, the rat had to approach the reward port on the opposite side of the box and emit two dry licks to the spout. On rewarded trials, the next three licks were followed by a drop of sucrose, with one drop per lick. In contrast, no sucrose was delivered on unrewarded trials, and the rat was required to deliver at least two additional dry licks (four in total). This requirement was essential to match oromotor responses to rewarded and unrewarded trials. Regardless of the outcome, the trial concluded 0.5 seconds after the last lick. Therefore, without external cues to guide their

choices or warn them about changing reward probabilities, the rats solely relied on the absence or presence of sucrose to inform their decisions.

In our task, we found that rats exhibited dynamic and adaptive behavior over a single session, which is the ability to change one's behavior in response to new information or changes in the environment. Rats were more likely to choose the lever that had previously rewarded them. The more likely a lever was to give a reward, the more likely the rats chose it (**Figure 1C**).

To model the rat's choice probability, we used Q-learning, a value-based reinforcement learning model. This approach uses the subject-specific history of rewards and actions to estimate action values, commonly called Q-values. In the rat's current state, the Q-value associated with an action reflects the expected value of taking that action. The increased Q-value of the left lever correlates with the higher likelihood of the rat choosing that option. The relationship between the choice probability and the delta of Q-values further improved after electrode implantation (as indicated by comparing the dashed black sigmoid curve with the solid green sigmoid curve in **Figure 1D**). Thus, electrode implantation did not impair the task's performance; rats kept improving. In **Figure 1E,** there was no significant preference (bias) towards either the left or right lever selection in the blocks highly rewarded leftward or rightward. This was supported by the non-significant t-tests (Unpaired t-test, 50:10 vs. 10:50 t-value $_{(246)}$ = 0.44, p-value=0.66; 90:50 vs. 50:90, t-value $_{(171)}$ = 0.54, p-value=0.54; 90:10 vs. 10:90, t-value $_{(162)}$ = -0.5, p-value=0.61). This result implies that rats did not inherently prefer one lever over the other, suggesting that the likelihood of receiving a reward was mainly responsible for guiding decision-making.

We next investigated how the delivery or omission of rewards affected the rats' decision-making strategy. We found that the probability of choosing the same lever from the previous trial increased as the number of reinforcements for that lever increased (Kruskal-Wallis test; stay probability: $\chi^2_{(7,\ 1985)}$ = 636.43, p<0.0001; **Figure 1F** red). This is known as the win-stay strategy. In contrast, the probability of choosing a different lever increased as the number of reinforcements for that lever decreased (Kruskal-Wallis test, shift probability: $\chi^2_{(7,\ 1985)}$ = 6.36.98; p<0.0001; **Figure 1F** blue). This is known as the lose-shift strategy. These findings suggest the rats were sensitive to reward history and employed a win-stay, lose-shift strategy in their decision-making process.

Following each block transition, the reward probabilities changed without warning, forcing the rat to update its beliefs about the current reward probability state in the block.[31] This led to an exploration phase, where the rat had to gather information about the potential value of different actions based on previous experience. This initial part of the block, arbitrarily defined as the first half within a block (25 trials), was named the Early phase. Once the rat had accumulated enough information, it transitioned to the exploitation phase, during which it could determine the most profitable action and consequently choose the higher-rewarding lever more often. This latter part of the block was designated as the Late phase (the second half of trials in the block). As expected, in the Early phase, rats initially exhibited a lower response rate toward the higher-rewarding lever. In contrast, in the Late phase, they consistently chose this lever more frequently, with the occurrence of this behavior increasing from 50:10 to blocks 90:50 and 90:10, where the disparity between the reward probabilities was greater, and the latter two blocks were more heavily rewarded (**Figure 1G**).

**Figure 1H** shows the probability of choosing the left lever across blocks. Note that rats chose the left lever more often in blocks where it had the higher reward probability (green) and chose the left lever less often in blocks where the higher reward was assigned to the right lever (purple; Kruskal-Wallis test, p (Left-Lever): $\chi 2_{(8,\ 1516)} = 361.14$, $p<0.0001$). In blocks with equal reward probability (gray blocks), rats exhibited a bias, responding more to one lever than the other. However, across subjects, this effect canceled each other out. Interestingly, in the block with the lowest reward rate (**Figure 1H**, see 10:10), rats were more likely to explore and constantly visit both levers than in the 50:50 and 90:90 blocks.

We also noted a proportional increase in rewards obtained to the highest reward probability in each block (**Figure 1I**). This observation was accurate not only for blocks highly rewarded leftward or rightward but also across the other blocks (**Figure 1J**; Kruskal-Wallis test, Reward rate: $\chi^2_{(8,\ 1516)} = 1034.88$, $p<0.0001$). Our findings indicate that rats could detect and adapt their choices in response to an environment characterized by nonstationary reward probabilities.

We finally investigated how the delivery or omission of reward modulated the licking behavior. Notably, our findings indicate that the number of licks given by the rat in the reward port was larger for rewarded trials compared to those where reward was omitted (Unpaired t-test, $t_{(79899)} = 347.5$; $p <0.001$; **Figure 1K**). In 97% of the rewarded trials (n = 47,007), the rats emitted at least 9 licks (see the dashed red vertical line in **Figure 1L**). In contrast, in approximately 98% of the trials where the reward was omitted (n = 32,894), they emitted (as

requested) at least four dry licks (see the dashed blue vertical line in **Figure 1L**). Thus, rats rapidly detected reward omission, as measured by their rapid cessation of rhythmic licking after four dry licks on unrewarded trials. This result suggests rats discriminate sucrose delivery from omission within a single lick,[32] highlighting the rats' remarkable ability to learn and adapt to changing reward probabilities. They tracked their previous choices, displayed high sensitivity to reward outcomes, and adjusted their choices while updating their internal beliefs. Importantly, this adaptive behavior solely relied upon the absence or presence of reward.

# Electrophysiology

## Single neuron analysis:

### Individual neurons encode choice, reward outcome, and rhythmic licking

A total of 1507 single units were recorded in the anterior and dorsomedial part of the NAc Shell region (NAcSh; **Figure 2 Supplementary 1**), while rats performed a self-guided probabilistic choice task. A time-warping algorithm was employed to realign the highly variable time intervals between epochs,[33] accounting for the variability introduced by the rat's pace. This process enabled the re-alignment of neuronal responses in the NAcSh. Specifically, the following task epochs were defined as: **Baseline**, one second before the rat's head entry into the central port and the onset of auditory stimulus. **Go**, the time interval between the onset of the auditory stimulus and the rat's head exiting the central port. **Choice**, the time interval between the rat's head exit and the lever press. **Anticipation**, the time interval between the lever press and the first lick delivered into the spout in the reward port. **Outcome**, the inter-lick intervals between the first and fifth lick given in the reward port. In this study, the employed time-warping algorithm linearly scaled the time interval for each epoch on a trial-by-trial basis, aiming to approximate the median interval observed across trials (see **Figure 2. Supplementary 2**). Applying this preprocessing to neuronal data allowed us to re-align the temporal structure of epochs, minimizing time lag variability

between trials and unveiling the neural activity across the entire trial structure (see **Figure 2. Supplementary 3**).

We next explored whether individual neurons had significantly different firing rate distributions for choice (left vs. right lever) or outcome (reward delivered vs. reward omitted), using the receiver operating characteristic (ROC) curves. Our data showed significant neuronal responses of the NAcSh associated with choice and outcome variables. For instance, a distinct subset of neurons ramped up their firing rate as the rat approached the reward port but was inhibited just before licking the spout. These neurons then exhibited a phasic activation during reward delivery but decreased activity during reward omission (**Figure 2A**, bottom panel, indicated by the red arrow). Accordingly, we labeled them "Outcome neurons." In addition, we recorded NAcSh neurons that exhibited selective responses when choosing either the left or right lever. A subset of neurons could encode two or more variables. As shown in **Figure 2B**, the same neuron was modulated by choice and outcome variables (top and bottom Peri-Stimulus Time Histogram (PSTH) panels, as indicated by the green and red arrows, respectively). This specific subset of neurons was termed Choice & Outcome neurons. Finally, we also identified neurons whose activity co-varied with rhythmic licking behavior (**Figure 2C**), and we referred to them as lick-coherent neurons. Neurons encoding the three variables choice, outcome, and lick-coherent were seen less often (more of this below).

From all neurons recorded, 31.1% (468 out of 1507) of NAcSh neurons were modulated by reward outcome; either they responded to delivery (15.6%) or omission (15.5%). NAcSh neurons also exhibited neuronal responses by reward outcomes, particularly during the Anticipation and Outcome epochs (with a minimal significant duration of 200 ms bins). Notably, we found the recruitment of omission-selective neurons during the Anticipation epoch before the actual omission of reward (depicted by the blue line and marked with a black arrow in **Figure 2D**). This implies a potential role of these neurons in expecting and predicting the omission of reward.

Similarly, NAcSh neurons exhibited responses related to the choice (400 ms minimal significant bin), spanning the Choice and Anticipatory epochs. A subpopulation of choice-selective neurons anticipated the rat's decision regarding which lever to press. Most of these neurons were recruited after the Go signal that preceded the lever press. Interestingly, a few

neurons carried choice-related information even during the Baseline epoch (i.e., before the trial began, indicated by a black arrow in **Figure 2E**). We propose that these anticipatory responses might reflect the subjects' probability of making a forthcoming choice.[34] The significant choice probability suggests that the neuronal activity provides information for predicting the subsequent behavioral response of the rat, such as pressing the left or right lever. We observed a higher number of right-selective than left-selective neurons. This is likely because recordings were obtained primarily from the right hemisphere (12 subjects, 1186 neurons) compared to the left (5 subjects, 321 neurons). These results highlight the capacity of NAcSh activity to guide decision-making, even without external sensory information, by integrating and predicting actions and their outcomes.

Outcomes modulated 31.1% of NAcSh neurons (468 out of 1507). Specifically, 10.9% NAcSh neurons (n = 164) were activated during reward delivery, while 9.5% (n=143) responded to reward omission (**Figure 2F**, Delivery and Omission). In addition, NAcSh neurons exhibited a significant ROC for reward delivery and subsequently for omission, or vice versa. While these neurons typically respond to a particular outcome, they could display distinct temporal activations. For instance, a neuron that exhibited a significant modulation for both reward delivery and omission but with a longer duration during the reward delivery activation was then designated as "Delivery preferred." We identified 4.7% of neurons (n=70) as Delivery preferred, while 6% (n=91) were classified as Omission preferred (**Figure 2F**, Outcome pie chart). This observation indicates that a comparable proportion of neurons responded to reward delivery and omission, suggesting that rats can use information from both outcomes to guide their choices equally.

The choice variable was encoded by 37.3% of neurons (562 out of 1507). Meanwhile, 8.6% of NAcSh neurons (n=130) were consistently activated for the left-lever choice, while 12.1% (n=183) responded to the right-lever choice. As in the outcome neurons, we also found neurons significantly activated by both left and right choices but at different time intervals. 7.4% NAcSh neurons (n=111) were left-preferred and 9.2% (n=138) were right-preferred (**Figure 2F**, Choice pie chart).

Regarding the lick-coherent neurons, 23.8% of neurons (358 out of 1507) exhibited significant coherence with rhythmic licking. Most of these neurons were lick-coherent either only in delivery trials (14%, n=211) or in both delivery and omission trials (D/O Coherent; 9.6%, n=144), while only a few neurons (0.2%, n=3) were coherent only in omission trials (**Figure 2F**, see Coherence pie chart).

We observed that NAcSh neurons were selective to one or more variables, with major overlap found in neurons modulated by outcome and choice (11.5%, n=173), followed by a small subpopulation modulated by all three variables (5.4%, n=81; **Figure 2G**). Conversely, a subset of lick-coherent neurons also conveyed information related to Outcome and Choice. For instance, 5.2% of neurons (n=78) were modulated by outcome and licking behavior, while 4.4% (n=67) were modulated by choice and licking behavior (see **Figure 2 Supplementary 3C**). This result suggests that these lick-coherent neurons were more complex than those that conveyed pure oromotor responses (**Figure 2 Supplementary 3C**). Finally, only 7.8% of neurons (n=119) were not responsive to any of these variables. These findings indicate that most NAcSh neurons (92.2%, n=1388) were engaged in the task, multiplexing two or more variables. In summary, the recruitment of outcome, choice, and lick-coherent neurons provides valuable insights into how single neurons in the NAcSh encode consummatory and reward-related information to guide decision-making.

## Population analysis:

We used two strategies to investigate the population activity of the NAcSh. First, we sorted the activity of neurons in rewarded trials according to their peak and then used the same sorting index during reward omission. This approach allowed us to visualize the sequential activation of NAcSh neurons, tracing the brain wave of reward information throughout the trial and assessing the stability of this representation across reward outcomes. Second, Principal Component Analysis (PCA) was applied to find the neuronal trajectories of NAcSh population activity evoked by reward outcomes. This enabled us to identify the population activity associated with reward delivery and omission.

### Neurons' peak activity sequentially represents key task events following a brain wave of reward information

Initially, we uncovered the activation pattern of NAcSh neurons by normalizing and sorting them based on their maximum activity and then plotting heatmaps. We did this independently for rewarded trials (**Figure 3A**, left) and then, using the same ranking index, for omission trials (**Figure 3A**, right). These heatmaps allow us to visualize the brain wave throughout the trial. We observed gradual recruitment of NAcSh neurons as the trial

progressed through various epochs: baseline 4%, Go 8%, Choice 27%, Anticipatory 41%, and Outcome 20%. A similar trend in recruitment was observed for omission trials (**Figure 3B**). This result suggests that NAcSh neurons sequentially monitored the entire trial and that more and more neurons were gradually recruited, triggered by the Go signal up to the outcome. Importantly, these ranked peaks were not random, as they were preserved during omission trials (**Figure 3A**, right), with a notable exception for the Outcome epoch. This sequential activation of NAcSh suggests they track the information flow underlying reward-seeking behavior, even without external sensory cues, by integrating activity across time bins.

Next, we evaluated peak stability across epochs by examining the proportion of neurons changing their peak during reward delivery compared to reward omission trials. In **Figure 3C**, the diagonal of the confusion matrix illustrates neurons maintaining their peak within the same epoch, providing a clear view of neuronal stability. Conversely, off-diagonal elements indicate epochs to which neurons shifted their peak from delivery to omission. Overall, the Anticipation and Choice epochs displayed the highest stability, with 79.13% (493 out of 623) and 70.52% (287 out of 407) of neurons, respectively, preserving their peak within the same epoch. In contrast, in delivery trials, neurons with peak activity in the Baseline epoch shifted their peak to the Anticipation and Outcome epoch in the omission trials. Neurons initially peaking during the Go and Outcome epochs in delivery trials showed a distinct shift in omission trials; these neurons primarily transitioned toward the Anticipation epoch and secondarily toward the Choice epoch. These findings reveal that while a subset of NAcSh neurons sequentially represent key task events and dynamically change their peak between epochs, the overall population activity maintains a stable representation across every time bin throughout the trial.

### NAcSh activity diverged into two distinct neural state spaces, reflecting both outcomes

To visualize the population dynamics of neuronal activity in the NAcSh, we employed principal component analysis (PCA), a simple dimensionality reduction technique.[35] While PC1 exhibited some association with licking behavior, initially, it could not distinguish reward outcomes, suggesting its role may be more related to motor control than reward processing (**Figure 3D**). This suggests that within the NAcSh, licking behavior accounts for most of the

variance (29.89%). PC2 exhibited an inverse relationship with PC1. Thus, it also co-varied with rhythmic licking, albeit negatively. In contrast, approaching the reward port triggered distinctive anticipatory activity ramps in PC3 and PC4. PC3 gradually increased activity and then dipped pre-licking, while PC4 ramped up after the lever press, dipped during the approach, and activated during licking. These trajectories effectively distinguished reward outcomes: PC3 neurons were dominated by selective activity for delivery (**Figure 3D**), while PC4 favored omission. Therefore, outcome feedback emerged as the second most prominent variable at the neuronal population level in the NAcSh, highlighting its role in detecting reward outcomes and guiding decision-making.

NAcSh activity tracks reward processing in three PCA dimensions: **Figure 3E** reveals three key components (PC1, PC3, PC4) that effectively capture the dynamic activity of neurons in the NAcSh during reward anticipation, licking, and outcome processing. These components explain nearly half (47.58%) of the total variation in firing rates across neurons. The red line depicts the average neural trajectory for rewarded trials, while the blue line represents omission trials. During the baseline epoch, both trajectories remain centered on the origin (0, 0, 0), indicating a stable neural state. However, as the task progresses, the trajectories diverge significantly, suggesting a perturbation in neuronal activity. Following head entry, both trajectories deviate from zero, suggesting increased neural activity. Interestingly, they exhibit an initial “twist” (potentially a phantom oscillation because of data smoothing [36]) between the Go signal and the rat’s exit from the central port. In the Choice epoch, the trajectories undergo a second “twist” and move further away from the origin, reflecting increased engagement. This continues until the lever presses, marking the transition to the Anticipation epoch. Here, a significant shift occurs, pulling the trajectories towards the positive quadrant of PC3 and the negative quadrant of PC4. The licking component explains the upward shift: This dramatic upward shift coincides with the rat protruding its tongue towards the empty sipper around 200 milliseconds before contact (**Figures 3D and 3E**). This strongly suggests that PC1, identified as a licking component, plays a key role in driving this shift. Taking these distinct trajectories in reward and omission trials together, the study highlights the dynamic nature of NAcSh activity during reward-guided behavior. The three identified components provide a valuable framework for understanding how this brain region encodes and processes information related to anticipation, licking, and outcome evaluation.

**NAcSh activity tracks reward vs. licking in outcome epoch**

It is important to note that until the outcome epoch, the NAcSh population activity, as represented by PCA trajectories, followed similar paths throughout the task (**Figure 3E**). However, within the outcome epoch, a striking shift emerges. Here, the rhythmic licking of the empty sipper becomes tightly coupled to phase-locked oscillations in the neural population. Similar oscillations were recently found brain-wide.[37] Interestingly, these oscillations diverge into distinct state spaces for rewarded and unrewarded trials, effectively separating the processing of reward from the act of licking. To quantify this divergence, we calculated the Euclidean distance between each time bin and the origin of the PCA space (**Figure 3F**). This distance reflects the overall "perturbation" of the population activity from its baseline state. As illustrated in **Figure 3F**, the PCA trajectories progressively move away from the origin (stable state) after the Go signal. Positive values on the y-axis indicate a stronger perturbation, with distance steadily increasing through the Choice epoch. Notably, within the Anticipation epoch, the Euclidean distance peaks just before the first empty lick (tongue protrusion) in both conditions. This highlights the progressive recruitment of neurons as the animal expects and attempts to receive a reward, regardless of the actual outcome.

Reward delivery diverges neural trajectories: In the outcome epoch, PCA trajectories oscillate in sync with dry licks (until the 3$^{rd}$ lick). However, a crucial shift occurs when a reward is delivered (1$^{st}$ rewarded lick). Here, distinct neural state spaces emerge, separating rewarded and unrewarded trials (**Figure 3E**). To measure this divergence, we calculated the Euclidean distance between rewarded and unrewarded trajectories (**Figure 3G**). Initially, they follow similar paths (across epochs), but a slight divergence during the anticipation epoch hints at reward anticipation (see arrow). This divergence sharply increases after the first reward delivery (yellow dashed line), highlighting the separation of licking-related activity from the reward response itself. A video graphic representation of these NAcSh population trajectories is available in **Supplementary Video Figure 3**.

Our population analyses suggest that the NAcSh tracks task progression and distinguishes reward outcomes. NAcSh neurons were gradually recruited in a brain wave pattern, peaking during anticipation and dissipating during reward outcome. PCA trajectories further corroborated these findings and revealed a robust distinction in neural activity patterns between reward delivery and omission, highlighting the NAcSh's key role in encoding the anticipation and the outcome of actions.

## Ensemble analysis:

### The heterogeneity of NAcSh responses is organized into "functional" neuronal ensembles and meta-ensembles

Leveraging PCA, t-distributed stochastic neighbor embedding (t-SNE), and hierarchical clustering analysis, we clustered neurons with similar firing patterns into neuronal ensembles (**Figure 4A**). Within the t-SNE map, all NAcSh neurons recorded were plotted and color-coded according to the ensemble assigned by hierarchical cluster analysis. Thus, the population activity was decomposed into 14 neuronal ensembles or processing modules, initially named neuronal ensembles A1-A14 (**Figure 4A**, see the Ensemble outcome map, in which the letter "A" indicates that "All" 50 trials in the block were used to compute the heatmap and t-SNE analysis). When we identified a plausible "function" of the neuronal ensemble, we renamed it accordingly. These 14 ensembles were automatically determined using the elbow method. Then, they were subsequently grouped into four meta-ensembles through a manual process (color-coded contours in **Figure 4A**) based on their proposed behavioral function: The epoch-related meta-ensemble (black contour) comprising ensembles monitoring each task's epoch. Choice-related meta-ensemble (green): Activity selective of the left or right lever press. Outcome-related meta-ensemble (red): Activity associated with the reward delivery or omission. A lick-active meta-ensemble (blue) combines the two ensembles that exhibited increased activity during licking.

The *Epoch meta-ensemble* (black contour) is composed of five neuronal ensembles: A1-Go (n=117) is selective to the Go epoch, and thus we renamed as Go, and so on, A3 (purple, n=107) fires more during go and choice epochs, and A11-Lev (n=189) is an ensemble responsive to lever press. A13-Ant (n=75) is an ensemble selective to the anticipation epoch, and the A9 (n=122) ensemble we proposed is goal-directed because it exhibited a sustained activity across multiple epochs, from Go, Choice, and Anticipation epochs. Thus, it could monitor all the steps needed to obtain a reward. An Epoch (Ep) meta-ensemble could track all-important task stimuli and epochs.

The *Choice meta-ensemble* (green) is formed by four neuronal ensembles named A5-Ch (Choice, n=118), A4 (ensemble 4, n=81), A7-ChL (Choice Late, n=118), and A6 (ensemble 6, n=33). All these ensembles fired more during Choice epoch, but as noted below, they all have information about choosing the left or right lever.

*The Outcome meta-ensemble* (red) comprises three ensembles: A2 (ensemble 2, n=59), Om (Omission, n=193), and Del (Delivery, n=97). A2 and Om fired more on omission trials, while Del was selective in reward delivery.

*The Lick-Active meta-ensemble* has two ensembles, A12 (ensemble 12, n=118) and Coh (lick coherent, n=80). These were the only two ensembles that exhibited more activity during licking; thus, we called them Lick-Active (LA).

Next, we investigated the proportion of neurons carrying information about decision variables (ROC curve significant activity) as a function of each ensemble (**Figure 4B**). **Figure 4B** shows the fraction of neurons with significant outcome (upper panel), choice (middle panel), and lick-coherent activity (lower panel). Epoch meta-ensemble: relatively few neurons encoded outcome, choice, and licking variables, except for ensemble "Go." In contrast, in the Choice meta-ensemble, a larger fraction of neurons carried choice information and lick-coherent activity (notably ensemble A6). Ensembles Ch and A4, besides encoding choice, also respond more to reward omission (resembling individual neurons responding to both Choice&Outcome see **Figure 2B** and **Figure 2 Supplementary 2A**). Moreover, the outcome meta-ensemble had a higher prevalence of neurons associated with reward outcomes, with two distinct patterns found: Ensembles A2 and Om had more neurons firing for omission than delivery. In contrast, the ensemble "Del" had more neurons selective for reward delivery. Ensemble Om also exhibited anticipatory activity that predicts the absence of reward (see Om **Figure 4C** and blue line **Figure 4D**), but during reward delivery, it fired phasically after the first drop of sucrose. Thus, this ensemble can detect both reward omission & delivery.

Last, in the Lick-Active meta ensemble, only two ensembles (A12 and Coh) exhibited increased activity during licking. They contained some outcome/choice encoding neurons, but ensemble "Coh" was particularly enriched in lick-coherent neurons. Thus, Lick-related activity, while less widespread, is notably enriched in specific ensembles, whereas choice and outcome information are mainly distributed across neuronal ensembles.

**Figure 4C** displays the normalized activity of all recorded neurons, organized into 14 ensembles and 4 meta ensembles, using a color-coded heatmap. In the left panel, trials with reward delivery reveal the temporal patterns of ensemble activation during reward anticipation and reward receipt. In the right panel, trials with reward omission highlight ensemble responses to the absence of reward. It can now be seen more clearly how the

population activity can be grouped into various neuronal ensembles with similar activity profiles (**Figure 4C**). Each ensemble's distinct activity profile revealed its neural dynamics and potential function. **Figure 4D** further dissects ensemble activity patterns: Left panel, ensemble PSTH activity, sorted by outcome delivery (red) and omission (blue), reveals distinct patterns reflecting different roles in reward processing. In the right panel, the same ensemble activity, but now sorted by choice of the left (green) and right (purple) lever, suggests potential involvement in decision-making processes. Each ensemble's unique spatiotemporal signature provides insights into its potential function and contributes to our understanding of the dynamic neural mechanisms underlying motivated behavior and decision-making.

Finally, the population activity (**Figure 4E**), the average responses of all recorded neurons, demonstrates that the NAcSh is modulated across all task epochs starting from the Go signal, including a noticeable anticipatory peak preceding the onset of licking followed by an inhibition (but still above baseline level) of NAcSh activity induced by sucrose, supporting the function of NAcSh as a "sensory sentinel" (see Introduction and [6]). However, here we found for the first time that this global NAcSh inhibition is still above baseline and thus is a reduction from the strong activation observed on the anticipatory epochs, notably during the first tongue protrusion licking the empty sipper. These findings provide valuable insights into the specific contributions of different neuronal ensembles to various behavioral variables. In this organization scheme, each NAcSh ensemble was a building block, or module, dedicated to processing different aspects of reward-guided behavior.

## Network analysis

### Functional interactions of NAcSh network exhibited a small-world architecture with a heavy-tailed distribution

We employed a probabilistic graphical model, specifically a Markov random field network to analyze the interactions of individual NAcSh neurons (assessed through covariance) and then grouped them by neuronal ensembles. This allowed us to quantify their pair-wise correlations and characterize their interaction pattern during different task conditions and trial types. As revealed by probabilistic modeling, **Figure 5A** depicts the intricate architecture of activity covariance within the NAcSh network. Each node (circle) depicts an individual

neuron, while edges (lines connecting nodes) signify their connections. The degree, or number of "functional" connections, was depicted as the size of the circles. The mutual information (MI) was calculated pairwise, in which the 90$^{th}$ percentile of the distribution was the upper cutoff to consider a connection (**Figure 5 Supplementary 1A**, Mutual information adjacency matrix); otherwise, a connection was rejected (**Figure 5A**; see Methods). **Figure 5A** illustrates the organic topology of the network, in which each node was colored based on the neuronal ensemble to which it belongs. The Delivery network, the left panel, illustrates NAcSh interactions during reward delivery trials, and the right shows the Omission network. Regarding the degree of connectivity, both Delivery and Omission networks exhibited a small-world architecture with a heavy-tailed distribution of 'functional' connections (see **SUPPLEMENTAL INFORMATION**). That is, most nodes (neurons) have a few connections (small circles), and only a few nodes are rich in connections (larger circles) and thus act as hubs (neurons particularly well connected). We can also see that hubs are present in the fourteen NAcSh ensembles but not in the same proportion (more below). Also, the average shortest path distance between nodes is much smaller than what is expected for regular or random graphs of similar dimensions, evidencing network shortcuts. These shortcuts are characteristic of certain functional networks as they allow for faster responses and the system's adaptability and dynamic robustness.

Next, we quantified some topological properties of the neuronal networks. We found a similar number of nodes containing at least one connection (N=1337 vs. N=1358), whereas the number of connections was similar in the Delivery and Omission networks (L=112799 vs. L=112800). We then evaluated the network properties regarding nodes. We found that the average shortest path length (SPL) between two randomly selected nodes in both networks was 2 steps (**Figure 5B** see dash lines). The network's diameter was ~4 steps (**Figure 5B**). This implies that moving from one node to another in the network can be achieved in only 2 steps. The network diameter measured by Φ was 4.23 steps in delivery and 3.99 for omission. As illustrated in **Figure 5A,** the Delivery network was slightly more expanded than the Omission network, which was more compact.

We next wondered whether the functional connections were homogeneously distributed among neuronal ensembles. Do neuronal ensembles rich in neuronal hubs exist? We observed the coexistence of weakly connected ensembles, for example, ensemble A9 (**Figure 5C**, indicated by left-shifted degree values for A9 green distribution), as well as three ensembles highly connected, Ch (pink), Coh (red), and Ant (blue) composed by many hubs

or neurons richly connected (**Figure 5C** upper panel). In the Omission network, the E9 ensemble was again weakly connected, but now four ensembles were richly connected (**Figure 5C** bottom panel, see Ch pink, Coh red, Ant blue, and A4 purple). Thus, unlike individual nodes, neuronal ensembles seem to follow a hierarchical organization, with some neuronal ensembles more connected than others. In particular, the Ch ensembles was always the most connected ensemble (see **Figure 8 supplementary 2**)

Then, we were interested in how the node degree affected the distance (d) between nodes and local connectivity (clustering coefficient). We observed that the distance between nodes decayed rapidly as the degree of connectivity increased, reflecting "short-cuts," while the clustering coefficient was less affected and did not decrease until nodes were highly connected (**Figure 5D**). In other words, the predominant short-cuts and the higher local connectivity (high cc) in the network facilitate the information flow. This interconnectivity between nodes indicates that the NAcSh activity follows a "small world" architecture [38], in which the clustering coefficient is high and the average path length is short, just as we found (**Figure 5D**). In this type of architecture, most nodes have a few connections and rare nodes are rich in connections and thus act as hubs.

Once we had evidence of rare neurons richly connected, we evaluated whether the degree distribution followed a long-tailed distribution (**SUPPLEMENTAL INFORMATION**). The slope of the power-law function fit was similar in the NAcSh network regardless of the outcome ($\gamma$=0.69 vs. $\gamma$=0.70, red and blue lines, respectively; **Figure 5E**). This result seems to support a scale-free-like property in both networks[39]. A more detailed hypothesis testing and goodness of fit calculation analysis can be found in **SUPPLEMENTAL INFORMATION**.

Another network feature concerning the degree distribution was the correlation degree, which reflects the connectivity pattern across the nodes (i.e., if hubs tend to connect or avoid each other). We observed homogeneous connectivity between all pairs of nodes, indicating that hubs neither preferentially connect nor avoid between them, although, for well-connected nodes $>10^2$, we saw a negative drop of $K_{nn}(k)$, suggesting that neuronal hubs tend to avoid connecting between them (**Figure 5F**).

**Sucrose delivery enhances the mutual information shared across the network**

Finally, we quantify the percentage of 'functional' connections across neuronal ensembles and the magnitude of the Mutual Information shared in the Delivery or Omission network. A

similar percentage of connections across meta-ensembles was observed for both outcomes (**Figure 5G**; Delivery vs. Omission, Epoch: 25.5% vs. 24.9%, Choice: 38.4% vs. 35%, Outcome: 19.8% vs. 22.8%, and Lick active: 16.3% vs. 17.3%). Surprisingly, and despite an equivalent proportion of connectivity in the meta-ensembles, the strength of mutual information shared in the NAcSh network was stronger in trials where the reward was delivered compared to those where it was omitted (Unpaired t-test: Delivery vs Omission, t-value $_{(225597)}$ = 161.08, $p<0.0001$; **Figure 5H**). This result suggests that sucrose delivery increased the mutual information shared across the network. This implies that the neuronal activity within the NAcSh network exhibited a more synchronized and coordinated response during the reward delivery than omission, suggesting that a sucrose drop enhances the flow of reward-related information.

**How neuronal ensembles interact during reward delivery and omission?**

Building on the revealed dense connectivity and small-world architecture of the NAcSh network, we aimed to characterize the strongest interactions between neuronal ensembles. We quantified the strength of "functional" connections (including self-loops) to identify ensembles with the highest percentage of interactions, representing the network's interaction backbone. While this approach does not provide directionality, it unveils the strongest communication landscape between neuronal ensembles. During reward delivery, the 1$^{st}$ strongest interactions of 13 out of 14 neuronal ensembles were strongly connected with the Choice ensemble (**Figure 6A**, see Ch larger black dots in the left panel). In contrast, during reward omission, the distribution of 1$^{st}$ and 2$^{nd}$ strongest interactions reshaped and exhibited a different interaction matrix (**Figure 6A** right panel). In reward omission, the Choice (Ch) - Omission (Om) and Coherent (Coh) ensembles formed an interaction core between them, and those three ensembles interact with the other neuronal ensembles (**Figure 6A**, right panel).

**Figure 6B** illustrates the interaction backbone for reward delivery. A radial interaction mode centered on the Ch ensemble emerged when isolating the 1$^{st}$ strongest interactions during reward delivery (**Figure 6B**, left panel). Surprisingly, most neuronal ensembles were coordinated by the Ch ensemble, except for the Go ensemble, whose nodes interacted more with nodes from the Coh ensemble. The Ch ensemble also exhibited the strongest interaction with itself, perhaps reflecting a role in facilitating the rat's integration of

information about choice and the outcome obtained, which is needed to update its decision process. This configuration suggests a potential computational aspect in the Choice ensemble resembling the function of the ventral striatum's action-value neurons.[40]

In contrast, during reward omission, the Ch and Om ensembles mainly coordinated the interaction backbone, followed by the Coh ensemble (**Figure 6B**, right panel). The Ch ensemble was more synchronized with the Om ensemble, and these two, in turn, interacted with the ensemble carrying lick-related activity, the Coh ensemble. Thus, our data suggest that the lick-coherent ensemble is poised to facilitate the integration and synchronization of relevant information in our task.[23] Collectively, these results unveiled the dynamic reshaping of interactions between neuronal ensembles in the NAcSh, depending on the reward outcome.

Having identified neuronal ensembles in the NAcSh and network properties using all 50 trials, we now investigated how their activity changed across the reinforcement learning within a block. We achieved this by repeating the ensemble analysis, focusing on neuronal activity from the Early or Late phases of the task.

**Neuronal ensembles are dynamic and reorganized by learning**

A notable feature of our behavioral task is the requirement for the rat to adapt its behavior and reset its reward expectations in each block transition. The rat solely relied on the delivery or omission of the reward as the external feedback to guide its choices. In the Early phase, the rat reset expectations and accumulated new evidence; thus, this is an exploratory phase. In contrast, in the Late phase, the rat became an expert and could firmly decide based on which lever was expected to yield the highest reward probability (exploitation phase; **Figures 1C and 1G**).

Given the importance of this adaptive foraging behavior, the present study examined how the neural activity patterns in the NAcSh were modulated during this reward-related learning process. We aimed to uncover the dynamic changes and identify specific neuronal ensembles involved in encoding and processing reward-related information and how they changed with learning. Once more, we employed t-SNE and hierarchical clustering to identify ensembles within the NAcSh during the Early and Late phases of the block, representing the first and last 25 trials, respectively.

Compared to the 14 neuronal ensembles identified using all trials, our automatic algorithm revealed only 12 and 10 distinct neuronal ensembles using half of the early and late phases

trials, respectively (**Figures 7A-B**). This gradual shrinking in the number of neuronal ensembles explains how each unique activity signature of ensembles consolidates and emerges from the population activity.

We used the Jaccard index (**Figure 7, supplementary 1**) to track individual neurons' transition from Early to Late phases among neuronal ensembles, which provided insights into how neuronal ensembles reorganized over learning. We observed that practically all neuronal ensembles reshape by removing neurons from and adding them to other neuronal ensembles, says that, overall, we noted that some, if not most, neuronal ensembles remained relatively stable across Early and Late phases, except for ensemble 7 (denoted as E7, "E" for neuronal ensembles found in "Early" phase) in the early phase, which was then dispersed between the delivery and omission (L-Del and L-Om, "L" stands for late phase) ensembles in the late phase. Conversely, the Coh ensemble was the most consistent neuronal ensemble across learning, with a few neurons adding in and out (**Figure 7B**).

In addition, ensembles E-Lev and E10 in the early phase were combined, giving rise to L-Lev formation in the Late phase (**Figure 7B**). Lick-coherent, Inhibitory, Choice, Delivery, and Omission ensembles (Coh, Inh, Ch, Del, and Om, respectively) exhibited distinct neuronal activity patterns that persisted throughout the block (**Figure 7C**). Interestingly, during the transition from Early to Late, most of the neuronal ensembles increased the number of neurons except for one ensemble designated as Choice-Late (L-ChL), which reduced the number of neurons in the ensemble during the transition. These findings highlight the stability of some and the dynamic nature of other neuronal ensembles and how their unique activity signature and neuronal composition emerged while solving a reinforcement learning task. Identifying and characterizing the dynamics of these neuronal ensembles might provide valuable insights into the functional interactions in the NAcSh network during a reinforcement learning task.

**Reinforcement learning reorganizes how neuronal ensembles interact**

A popular idea suggests that the subject's decisions in reward-related learning tasks rely on estimates linked to different actions.[41] In the later phase of the block, rats select the lever associated with the action with the highest estimated reward likelihood. This selection is guided by neural representations of these action-associated estimates, sometimes called action values within the striatum.[40] Here, we proposed that one possible neural

representation might be contained in the Ch ensemble, which displayed robust and sustained activation during the choice epoch. The Ch ensemble consistently maintained the largest connections within the NAcSh network across reward outcomes and reward-related learning from the Early to Late phase (see **Figure 8 Supplementary 2**).

Our next goal was to characterize how the strongest interactions among neuronal ensembles were reshaped by learning. During the Early phase, the network representation of reward delivery in the NAcSh exhibited how the modular organization in the NAcSh emerged, and it was primarily synchronized with the early Choice neuronal ensemble (E-Ch) as its 1$^{st}$ strongest interactions. E-Ch ensemble formed a "closed-loop" with the neuronal ensemble 3 (**E3**, **Figure 8A**, top left panel). Interestingly, the 2$^{nd}$ strongest interactions (below) converged mainly between the E-Om and E-ChL neuronal ensemble, which interact with E-Coh (**Figure 8A**, bottom left panel). Remarkably, during the Late phase, the radial and closed-loop organization was further taking form and was present in the 1$^{st}$ and 2$^{nd}$ strongest interactions. The 1$^{st}$ strongest interactions were organized radially with L-Ch, and this ensemble formed a closed-loop with L3. The radial organization persisted in the 2$^{nd}$ strongest interactions and was mainly coordinated with L3. In addition, L3 and L-Ch formed self-loops. Thus, L-Ch and L3 (both members of the Choice Meta-Ensemble) appeared to synchronize the NAcSh network (**Figure 8A**, right panel).

Now, the network representation of reward omission in the NAcSh, during the Early phase, the 1$^{st}$ strongest interactions appeared to be organized hierarchically, primarily by the core interaction between E-Ch and E-Om, the latter neuronal ensemble coordinated with the other Outcome ensembles (E-Del and E7), E-Ant, and E-ChL. Secondarily, E-Om ensemble was synchronized with the E-Coh, resembling a bridge with most of the neuronal ensembles within the Epoch Meta-Ensemble. For the 2$^{nd}$ strongest interactions, the E-ChL acted as the bridge, with a strong connection to E-Coh, E-Om, and E-Ch. During the Late phase, the configuration of the 1$^{st}$ strongest interactions revealed three well-defined aspects. First, the L-Ch was consistently the backbone of the structured interactions in the NAcSh network and exhibited a strong self-interaction. Second, L-Lev acted as a connector, linking L-Ch with L-Coh, akin to an interface, facilitating communication between decision variables and licking behavior. Third, L-Del ensemble emerged as an isolated module. An interesting configuration emerges in the 2$^{nd}$ strongest interactions. The Choice and Epoch Meta-Ensembles were coordinated by L3 ensemble. On the other hand, L-Coh formed a "triangle"

with L-Go and L-Om and synchronized with the rest of the ensembles in the network. L3 and L-Coh linked these “two parts” of the network.

All in all, our data reveal not only distinct interaction patterns among neuronal ensembles during reward delivery and omission, but also a progressive reshaping of these interactions driven by learning.

**Reinforcement learning enhances synchronization (mutual information) during reward delivery and diminishes it by reward omission.**

Finally, we evaluated how the strength connectivity of the NAcSh network was affected by the reinforcement learning. In trials in which the reward was delivered, the mutual information increased during the reward-related learning (MI Reward delivery, Unpaired t-test: Early vs Late, t-value $_{(225,596)}$ = -70.92, p<0.0001; **Figure 8C**, left panel). In trials in which the reward was omitted, the mutual information was not affected uniformly across the neuronal ensembles. Only the neuronal ensemble 3 and ChL increased (MI Reward omission, Unpaired t-test: E3 vs L3, t-value $_{(42,281)}$ = -6.4, p<0.0001; E-ChL vs L-ChL, t-value $_{(38,850)}$ = -15.29, p<0.0001). However, the remaining ensembles decreased significantly (MI Reward omission, Unpaired t-test: E-Go vs. L-Go, t-value $_{(36,235)}$ = 8.31, p<0.0001; E2 vs. L2, t-value $_{(6,125)}$ = 3.57, p=0.0004; E-Lev vs. L-Lev, t-value $_{(36,090)}$ = 8.09, p<0.0001; E-Ant vs. L-Ant, t-value $_{(30,764)}$ = 15.71, p<0.0001; E-Ch vs. L-Ch, t-value $_{(61,856)}$ = 17.87, p<0.0001; E-Om vs. L-Om, t-value $_{(45,809)}$ = 8.65, p<0.0001; E-Del vs. L-Del, t-value $_{(35,152)}$ = 8.44, p<0.0001; E-Coh vs. L-Coh, t-value $_{(50,567)}$ = 9.65, p<0.0001). These results indicate an opposite effect on mutual information shared in the NAcSh ensemble during reward-related learning and outcome. Collectively, the NAcSh network reshapes the strongest interaction, and sucrose enhanced network synchrony, whereas reward omission decreased it. The reinforcement learning progressively refined these effects (See Graphical Abstract for a summary).

## Discussion

Our study reveals how the NAcSh neurons encode reward information through dynamic neuronal ensembles. As rats adapt their choices based on the reward likelihood of a sweet reward (**Figure 1**), these neuronal ensembles progressively reorganize, with individual neurons dropping in and out, enhancing their communication strength during reward

delivery. Importantly, individual neurons encode not only reward outcomes but also the rats' choices and licking behavior, with some neurons representing multiple variables at different times. Thus, it is clear that the NAcSh does more than track reward.[3] This neuronal activity unfolds like a brain wave peaking during anticipation and dissipating upon outcome delivery, potentially reflecting a global brain wave passing through the NAcSh.[42] Using PCA, we confirmed the progressive recruitment of NAcSh neurons, culminating in a crescendo-like pattern upon reward delivery. Similar to the rats' behavior, PCA trajectories precisely distinguish reward from omission after a single drop of sucrose. Further analysis identified fourteen distinct ensembles organized into four functional meta-ensembles: Epoch, Choice, Outcome (Delivery-Omission), and Lick-Active (LA). This modular organization suggests that each ensemble acts as a distinct module for processing distinct features of reward-seeking behavior. Delving deeper, we discovered a small-world, long-tailed network architecture within the NAcSh. This architecture optimizes information flow, with most neurons having few connections and a few hubs holding numerous "functional" connections (correlations). Notably, NAcSh neurons share more mutual information with the network during reward delivery, signifying increased communication (or coordination). Furthermore, ensemble interactions dynamically reshape during outcome delivery and omission. Interestingly, as the rat becomes more expert through reinforcement learning, ensembles merge (shrink in number), and network connections become highly synchronized, suggesting optimized processing during reward delivery. These findings offer a novel perspective on reward processing in the NAcSh. Long-tailed network architecture and dynamic ensembles that adapt as rats learn to obtain rewards suggest that NAcSh activity guides decision-making even without external cues, potentially predicting the outcome of their actions.

**NAcSh Neurons Encode Choice, Outcome, and Licking:**

Individual NAcSh neurons exhibited diverse activity patterns, representing the rat's choice (lever selection), reward outcomes (delivery or omission), and rhythmic licking (**Figure 2**). Some neurons encoded multiple variables at different times, suggesting they could convey multiple messages. Choice-selective neurons could participate in action selection, and some even predicted lever choice before the trial began (in baseline). These findings align with the role of the NAc in action monitoring,[43] action selection,[40] as well as predicting and guiding actions.

**Outcome Encoding in NAcSh:**

We observed both delivery-selective and omission-selective neurons (**Figure 2 Supplementary 3A**), implying potential learning from both outcomes. Importantly, licking differences are unlikely explanations because we forced rats to deliver > 4 licks in both outcomes. Under this orofacial controlled condition, we confirmed Regina Carelli's findings of omission-responsive neurons[4] and found them equally represented in the NAcSh. PCA analysis further confirmed two outcome-related PCA components (**Figure 3**), PC3 (reward delivery) and PC4 (reward omission). Thus, outcome encoding was the second most important variable after licking in the NAcSh, possibly encoding both rewarding and punishing signals. Considering that reward omission served as a punishment, and thus, it could be a mild form of aversive stimuli, our results support the idea that NAcSh encodes both appetitive and aversive stimuli.[17,44–46] A new finding was the discovery of some omission-selective neurons that predict the omission ("pessimist" neurons; see also Om ensemble **Fig. 4D**). This is even when no cue signals whether the rat would receive a reward. Thus, the NAcSh could also participate in an "internal" prediction of reward omission.

**Lick-Coherent Neurons:**

We identified lick-coherent neurons previously reported in our studies.[12,23,24] While comprising only 23.8% of neurons, they exhibited the most robust response, residing in PC1 and PC2 components. This makes licking the most represented variable in the NAcSh. Notably, our findings resonate with recent studies in head-fixed mice and high-density neuropixel recordings that found a brain-wide representation of reward outcome.[40,47–49] They conclude it might partly or primarily reflect the licking movements required for reward consumption rather than the hedonic aspects of reward.[37] Here, we confirmed the important role of licking, but thanks to our task design, we could also clearly distinguish reward signals from those encoding licking itself.

**Reward information flows in a global brain wave.**

We show that the peak activity of NAcSh neurons sequentially represents the entire trial; these neural sequences were not random, and it has been hypothesized that they could help keep tracking the passage of time.[50] If this is the case, this population sequence could keep the timing of goal-directed behaviors needed to obtain a reward. Alternatively, this potentially implies that reward information flows in a spatiotemporal and global traveling wave(s) that pass throughout the accumbens. Recent studies have confirmed that representations of

choice and reward outcomes were widely distributed across the brain.[37,47–49] The same global activation was observed when rodents consumed sucrose.[23,42,51] In particular, the ventral striatum, including the NAcSh, has more neurons recruited than other striatal regions by the outcome feedback.[40,42] Our PCA analysis also confirmed the gradual recruitment of NAcSh neurons triggered by the Go tone peaking up to reward anticipation, which dissipates during the outcome port, in which information flow clearly distinguishes orofacial responses from reward signals.

We note that disentangling licking from reward is a challenging task.[37] This is because previous studies reported anticipatory and suppressed activity in the NAcSh during freely consummatory behavior,[12,14] supporting the notion that NAcSh inhibition gates food intake.[6,15,16] However, separating reward signals from licking itself is difficult because of their close temporal association. For example, analyzing NAcSh activity under freely licking sucrose intake confounds reward with oromotor activity, hindering signal separation.[12,24,37,45] Here, we propose a novel approach to delaying sucrose delivery (after two empty licks) and enforcing dry licks during reward omission trials. This effectively isolated reward signals from orofacial responses, revealing them as brief, phasic peaks within one or two licks (also see Del Ensemble **Figure 4D**).

**NAcSh as a hedonic feeding sentinel**

Population neuronal activity revealed NAcSh modulation across all task epochs (**Figure 4E**), initiating with the Go signal, followed by an anticipatory peak before licking and a licking-induced inhibition. These findings support the proposed “sensory sentinel” function of NAcSh.[6] Notably, we observed for the first time that this global NAcSh inhibition, which still exceeds the baseline, represents a reduction from the robust activation seen during anticipatory epochs and first tongue protrusion. Our subsequent ensemble analysis further confirmed that, from 14 neuronal ensembles identified, only two ensembles were active during licking (see the Lick Active meta-ensemble). These results confirmed a hallmark feature of the NAc (see Introduction). That is the suppression during palatable food or sucrose intake.[10–12,14,45] Interestingly, the anticipatory peak and the consummatory inhibition in the NAcSh were observed even when the first two licks were empty or during omission trials, in which all licks were, in fact, empty. Thus, this demonstrates for the first time that it is not sucrose or the palatable food *per se* that inhibits NAcSh activity; instead, it is pure licking. We acknowledge that this inhibition occurs when the rat expects to obtain sucrose.

While our single sucrose concentration used as a reward precludes assessing neurons' ability to encode hedonic value, previous studies suggest NAcSh encodes both palatability and sucrose concentration.[11,18,22,24] Therefore, taste hedonic signals in our recordings might be obscured. However, we note that sucrose-responsive neurons rarely respond equally to all three sucrose drops, suggesting they do not faithfully encode taste information or they exhibit taste habituation.[52]

**Neuronal ensembles in the NAcSh are dynamic and serve as distinct modules for reward-related information**

Cyriel Pennartz, in a review, was the first to hypothesize that the NAc is organized into ensembles and proposed that NAc functions are based on neuronal ensembles of active and inactive neurons.[53] Patricio O'Donnell further suggested that ensembles must be adaptable to a changing complex environment depending on the spatiotemporal distribution of limbic inputs on active and inactive neurons, reflecting their up and down bistable membrane potential of medium spiny neurons.[44] He further elaborates that these neuronal ensembles could be established and reinforced by several glutamatergic limbic inputs, yielding dynamic ensembles of NAc involved in reward and cognitive functions.[44] In this view, reward (and aversion) are the driving forces reinforcing (or punishing) the formation of dynamic neuronal ensembles. Building on the seminal work of Pennartz et al. (1994) and O'Donnell (1999), we present compelling evidence for dynamic NAcSh ensembles. Our findings align with the notion that ensembles are shaped by rewarding glutamatergic limbic inputs[16,21] and reward outcomes, reinforcing their role in reward and cognitive functions. Extending these classic ideas, we reveal the dynamism of NAcSh ensembles, including their network properties and how their connectivity interacts with reward outcomes. This underscores the crucial role of dynamic ensembles in understanding NAcSh function.[18,22]

Our recordings identified several important neuronal ensembles encoding reward-related information. We propose that the epoch meta-ensemble coding specific task epochs might reveal the current state or the entire trial structure. The Coh ensemble would also be important in defining the outcome epoch (more below). The Ch ensemble was highly informative and not only predicted lever choice (left or right) but also exhibited the strongest connections with other ensembles, suggesting its central role in coordinating reward delivery interactions.

Reward-related ensembles: We identified several ensembles firing selectively for reward delivery, notably the Del ensemble. In contrast, other ensembles like Om (predictive but also firing during delivery), E6, E4, Ch, and E2 showed more robust activation for reward omission. This cornucopia of neuronal ensembles revealed the significant heterogeneity of NAcSh responses and supported the idea that neuronal ensembles are distinct modules specialized in processing specific reward-related information.[27]

**On the complexity of the task and the number of neuronal ensembles**

Our study employed a self-guided probabilistic choice task, known for its higher cognitive demands, as it requires rats to generate internal representations of their environment.[54] Compared to simpler tasks like freely licking[12] or brief-access taste tasks,[22] this paradigm revealed a striking expansion in the number of active NAcSh ensembles—up to 14 distinct modules. This suggests a direct link between task complexity and the number of ensembles recruited, implying that more information processing demands cause the activation of additional dedicated modules within the NAcSh. This finding strengthens the hypothesis that NAcSh neuronal ensembles function as specialized units or modules, dynamically adapting their number and composition in response to the behavioral context and information needs of specific reward-guided behaviors.[44]

**NAcSh covariances follow a small-world network architecture with a heavy-tailed degree distribution.**

Our findings align with Watts and Strogatz's original "small-world" proposal for the brain's architecture.[38] This likely arises from a combination of random connection pruning, synaptic strength rearrangement, and preferential attachment dynamics.[39] Such mechanisms resemble Hebbian self-organization, suggesting that heavy-tailed connectivity might be a general feature of neural networks.[55] Recent network analyses of the synaptic brain connectome across various organisms (from fly to mouse) support this, revealing heavy-tailed distributions. Moreover, calcium activity within the visual cortex neurons of a single mouse seeing natural images also exhibits this long tail, "free-scale" property,[55] echoing our observations in the NAcSh activity. Strikingly, our findings on small-world architecture with heavy-tailed properties in the NAcSh of the ventral striatum mirrored those observed in brain slices from the dorsal striatum.[56] Thus, the dorsal and ventral striatum exhibited an

asymmetrical tail in the degree distribution of “functional” connections between co-active striatal neurons, consistent with a long tail distribution (**SUPPLEMENTAL INFORMATION**).[60]

### Sucrose reward enhanced information flow within the network

Unlike reward omission, neuronal ensembles in reward delivery trials exhibited significantly higher mutual information, indicating stronger synchronization and coordination among network nodes. This suggests a facilitated flow of reward information across NAcSh ensembles, enabling effective processing and integration. Conversely, mutual information was lower during reward omission, reflecting reduced network-wide communication. Notably, both increases and decreases in mutual information with reward delivery and omission progressively improved across learning within a block (as shown in the Graphical Abstract), highlighting the dynamic nature of information processing within the network.

### Licking orchestrate network synchronization

Unexpectedly, licking emerged as a key player in orchestrating network synchronicity. Across various analyses (PCA, ensemble, network), the Lick-coherent ensemble (Coh) consistently demonstrated a central role. Regardless of the reward outcome, the Coh ensemble displayed strong interconnectivity and high mutual information (**Figure 5H**), suggesting its crucial function in coordinating neuronal activity. This finding aligns with the proposal that rhythmic licking acts as an internal clock, synchronizing brain-wide activity and information flow.[23] Coh ensemble’s high mutual information is potentially linked to its restricted activity during the outcome epoch. This unique firing pattern may allow it to convey more specific information about the outcome than other ensembles.[23,51,57]

### Reinforcement learning reconfigures and consolidates neuronal ensembles

We found that the composition of our neuronal ensembles changed as a function of learning within blocks, with neurons removed from and added to another ensemble. Our t-SNE and hierarchical clustering analysis revealed 12 and 10 early and late phases ensembles. This shrinkage in ensemble number most likely reflects consolidation [58] and a larger similitude of

firing patterns among neurons that emerge in the late phase of the block. This reorganization was also observed in the strongest interactions between ensembles from the Early and Late phases. These results highlight the strong dynamism of neuronal ensembles in the NAcSh and its potential involvement in higher cognitive functions, as initially proposed by Patricio O'Donnell.[44]

**Implications for neuroscience: Can we catch a wave?**

We propose that the next challenge in neuroscience is to map how genetically defined cell types correspond to "functional" neuronal ensembles based on neural activity. In a tour de force, research led by the Allen Brain Institute unveiled 5322 distinct cell types in the mouse brain.[59] This raises the question of which cell types compose the "functional" ensembles. Perhaps each neuronal ensemble is composed of one and only one distinct cell type that exhibits a particular activity signature and function.[60] Alternatively, does each "functional" ensemble comprise various cell types with similar firing patterns? Further studies are needed to answer these questions, but evidence supports both hypotheses.[12,17,18,22,60] However, given that a single cell type within a brain region may play multiple functions, it suggests that various cell types may be distributed across "functional" neuronal ensembles.

Recent techniques such as Fos, Arc TRAP [61], Cal-Light [62], or mArc [63] with a shorter time window promise to catch neurons activated by a stimulus or behavior, and thus, at first sight, it will be ideal for catching "functional" ensembles or even an engram that can induce a percept [64] then reactivate or erase that memory.[65] However, suppose information flows in a brain wave, as our data and those from many other labs suggest.[49,51] In that case, trapping these "functional" ensembles is a real challenge. For example, suppose we want to catch neurons based on activity and time window (e.g., within the Choice epoch). In that case, catching selectively, the Ch ensemble will also be trapped with all other neuronal ensembles firing more during this epoch (i.e., A3, Lev, A4, ChL, A6, A9) and with active neurons on those trials (e.g., neurons responding to left-right choice, rewarded or not). Moreover, suppose neuronal ensembles are dynamic,[66] and their neuronal composition changes with learning, as we and many others suggest.[58] In that case, some neurons trapped will also change depending on the learning progress of the task and whether the trial was rewarded. Our results predict that these impressive new techniques still will not be sufficient to dissect our "functional" ensembles with the precision needed.

### What is the cellular identity of neuronal hubs?

Our network analysis revealed a “rich get richer” connectivity pattern, resembling networks with preferential attachment.[39] This architecture facilitates rapid information transfer and is resilient to random deletions (removal of weakly connected neurons). However, it becomes vulnerable when explicitly targeting the highly connected “hubs.” [67] This raises intriguing questions: who are these hubs? Are they specific cell types, like interneurons? [26,67] If so, they constitute less than 10% of NAcSh neurons.[68] Studies in brain slides from the dorsal striatum suggest that most hub neurons are, in fact, interneurons, fast-spiking, low threshold spiking, and cholinergic interneurons.[67] Alternatively, could they be medium spiny neurons (MSNs, comprising ~90% of NAcSh neurons)? If so, what distinguishes hub MSNs from other nodes of MSN neurons? Future studies should explore these fascinating avenues to illuminate these critical network elements’ nature and functional significance.

### Caveats and limitations

Here, we used PCA and t-SNE followed by hierarchical clustering, to define “functional” neuronal ensembles. While different dimensionality reduction and clustering methods like UMAP,[69] CEBRA,[70] or graph theory[67] might yield slightly different ensemble numbers, the fundamental observation remains consistent: neuronal ensembles emerge[18,22] and play a crucial role in brain computations, regardless of the specific mathematical approach.[27] Likewise, our correlation matrix leverages data from multiple sessions and rats, capturing the strength of “functional” connections related to task events in a kind of “meta-brain” of 17 animals performing the same behavioral task, and thus, it lacks individual-level detail because of restrictions imposed by current recording methods. Notably, even recent advancements like Neuropixels are primarily optimized for head-fixed setups,[37,47,49] but see [71]. We need novel high-density electrode technologies for freely moving rodents to record more neurons simultaneously within a single brain, unlocking a deeper understanding of individual neural dynamics under more naturalistic conditions than those provided by head-fixed setups.[72]

### Conclusion

NAcSh ensembles exhibit remarkable dynamism, progressively reorganizing throughout learning as individual neurons drop in and out, changing the composition of neuronal ensembles. Importantly, reward delivery enhances network synchronization, suggesting coordinated processing of reward information. Interestingly, the NAcSh network follows a heavy-tailed distribution, meaning a few highly connected neuronal hubs play a central role in information flow. This architecture facilitates the efficient transmission of reward signals within the network.

## STAR★Methods

### Subjects

Seventeen male Sprague-Dawley rats (300-350 g) were individually housed in a temperature-controlled room (22 ± 1 °C) with a 12-hour light/dark cycle (lights on at 06:00, off at 18:00). The rats were water-deprived and only had access to water for 30 minutes after each behavioral session. Chow food (PicoLab Rodent Diet 20, MO, USA) was always available *ad libitum*. The Animal Care and Use Committee of CINVESTAV approved all procedures.

### Apparatus

Training and recording were conducted in a standard operant conditioning box (30.5 x 24.1 x 21.0 cm; Med Associates Inc., VT, USA). The box had a front panel with a central V-shaped port with an infrared sensor to detect head entry. Two retractable levers were also placed on the sides of the central port. On the opposite wall was a V-shaped licking port with an infrared sensor to detect individual licks. This port contained a licking spout connected to a solenoid valve (Parker, OH, USA). The spout delivered a sucrose solution (10% w/v sucrose dissolved in distilled water). The volume of each drop was calibrated daily to ~10 µL using an independent air pressure system.[23]

### Behavioral task

A self-guided probabilistic choice task was used,[28,30] also known as a 2-armed bandit. The training period for the trial-and-error task lasted from 30 to 60 days and included a first pretraining phase (2 days to 2 weeks) before the rats progressed to the entire task.

### Pre-training phase

To complete the pretraining phase, rats had to complete at least 150 trials. Each trial began with the house lights illuminated, which instructed the rat to approach the central port and place its head for at least 100 ms. Once this fixed interval elapsed, an auditory stimulus was delivered, and a single lever was extended. The lever alternated between the left and right lever on consecutive trials (e.g., if the left lever was extended in the current trial, the right lever would be extended in the subsequent trial). After the lever was pressed, the rat had to approach the reward port on the opposite side of the box. The rat was required to deliver two dry licks to the spout, and three drops of sucrose were delivered in the subsequent three licks. All trials were rewarded in this pretraining phase, and sessions were limited to 30 minutes.

**Complete task**

In the complete version of the task, the rat was required to wait 250 ms before an auditory stimulus was presented. Both levers were then extended into the box. The auditory stimulus prompted the rat to perform a leftward or rightward lever press but did not provide guidance on which lever would deliver the highest reward. Reward probabilities were assigned to the left and right levers, covering all combinations of 10%, 50%, and 90%. The reward blocks changed every 50 trials without an explicit signal, and the order of blocks varied across sessions.

After the lever press, both levers were retracted simultaneously to prevent the possibility of a second press in a single trial. The rat then had to approach the reward port on the opposite side of the box and give two dry licks to the spout, regardless of the outcome. In rewarded trials (reward delivery), the spout delivered a drop of sucrose in the following three licks. In unrewarded trials (reward omission), the rat had to give two more dry licks to finish the trial (i.e., at least 4 dry licks). Trials ended 0.5 s after the last lick was given in the reward port. No external cues could influence the rat's choices aside from the outcomes of its actions.

**Surgery**

Animals were anesthetized with ketamine (70 mg/kg, i.p.) and xylazine (20 mg/kg, i.p.). They were then placed in a stereotaxic apparatus, where the skull was exposed, and craniotomies were made in the skull over the recording site. Two screws were set for holding, and a silver wire was soldered to the third screw. This served as an electrical reference for recordings and was inserted above the cerebellum. A custom-made 16-tungsten electrode array (35 µm diameter, each arranged in a 4 x 4, area = 1 mm²) was implanted into the NAcSh (AP: +1.2 mm, ML: ±1 mm from bregma; DV: -7.6 mm). The electrode array was stained with a

fluorescent dye (cell tracker CM-DiI dye, red excitation/emission spectra: 553/570 nm; ThermoFisher Scientific, MA, USA) just before it was inserted into the brain. Dental acrylic was applied to the screws, making a stable attachment to the skull and fastening the electrode array. The rats were given ketoprofen (45 mg/kg, i.p.) three days after the surgery and recovered for one week.

### Histology

Once electrophysiological signals disappeared, the animals were deeply anesthetized with an overdose of pentobarbital sodium (150 mg/kg, i.p.). They were then transcardially perfused with phosphate-buffered saline (PBS 1x) followed by 4% paraformaldehyde. Brains were extracted, stored for 1 day in 4% paraformaldehyde, and then transferred to a 30% sucrose/PBS solution. Brains were cut at 60 μm coronal slices. An epifluorescence microscope confirmed the trace and last position of the electrode array tips.

### Electrophysiological recordings

Recordings were made while rats performed the probabilistic choice task. Neural activity was amplified, digitized with a 64-channel PZ5 Neurodigitizer amplifier (Tucker-Davis Technologies, FL, USA), and then transmitted to the RZ2 BioAmp Processor (Tucker-Davis Technologies, FL, USA). Signals were sampled at 50 KHz (28 bits of resolution) and band-pass filtered from 0.5 to 8 KHz. The action potentials were identified online using voltage-time threshold windows. To be classified as a single-unit, action potential needed to have an amplitude above 50 μV, and their spikes had to exhibit an amplitude of 3.5-4 standard deviations above the mean signal of the channel. The spikes were exported (OpenBridge software, Tucker-Davis Technologies, FL, USA) and sorted using Offline Sorter Software (Plexon, TX, USA). Only time stamps from the offline-sorted waveforms were analyzed.

### Statistical analysis

All analyses were performed with MATLAB (Mathworks) and R (http://www.r-project.org/). All data are presented as mean ± SEM unless reported otherwise. For nonparametric tests, the Wilcoxon rank-sum test was used.

### Reinforcement learning model

Q-learning model was employed to predict the decision-making behavior (probability to choose left lever, p(Left-Lever)) by leveraging subject-specific reward history and action

sequences to estimate action value. Action-values were calculated on a trial-by-trial basis and updated to the following algorithm:

$$Q_i(t+1) = \begin{cases} Q_i(t) + \propto \left(R(t) - Q_i(t)\right), & if\ a(t) = i \\ Q_i(t), & if\ a(t) \neq i \end{cases}$$

Where $a(t)$ and $R(t)$ indicate the choice and reward in the current trial, respectively, $i$ is the action associated with the value (e.g., left or right lever), and $\propto$ is the learning rate. This model states that in two alternative task, the probability of choosing an action is a sigmoidal function of the difference in the action values:

$$p(Left - Lever) = \frac{1}{1 + e^{-\beta(Q_L(t) - Q_R(t))}}$$

Where β is a parameter that determines bias towards the action associated with the higher action value. The learning rate and bias parameters were estimated from the behavior for each session as in [73].

**Time-warping algorithm**

We determined 5 epochs in the task: 1 s before the head entry to the onset of the tone (Baseline), from the onset of the tone to head exit (Go), from the head exit to the press lever (Choice), from the press lever to the first lick (Anticipatory), and from the first to the fifth lick (Outcome). These epochs were aligned by implementing the time-warping algorithm, given the variability of the time intervals across trials. This algorithm was not applied to the Baseline epoch. The epochs were aligned by calculating the median of the time intervals (**Fig. 2 supplementary 1**) for Go = 0.35 s, Choice = 1.3 s, and Anticipatory = 1.9. The outcome epoch was aligned lick-by-lick. Thus, we calculated the Inter-Lick Interval (ILI) median = 0.14 s. The duration of an epoch was calculated in each trial:

$$\Delta E(t) = E_1(t) - E_2(t)$$

where $E_{1,2}(t)$ are the events which start and terminate the epoch, respectively, and $t$ is the trial. We defined an upper limit ($UL_{Epoch}$) for each epoch duration given its cumulative density function. If $\Delta E(t)$ exceeds its corresponding epoch limit, $\Delta E(t)$ was determined following this formula:

$$\Delta E(t) = E_2(t) - UL_{Epoch}$$

We considered these trials as delayed, avoiding the over-shrink epoch. To set the spike timing in the same interval of the duration of the epoch was given by:

$$\Delta\, spikes_{1:n}(t) \;=\; Spikes_{1:n}(t) - E_1(t)$$

Or when the epoch duration exceeds its upper limit:

$$\Delta\, spikes_{1:n}(t) \;=\; Spikes_{1:n}(t) - UL_{Epoch}$$

Where $Spikes_{1:n}(t)$ are the spike timestamps in the epoch. A range normalization of the epoch was calculated as follows:

$$\Delta spikes'_{1:n}(t) = \frac{\Delta spikes_{1:n}(t)}{\Delta E(t)}$$

$$0 < \Delta spikes'_{1:n}(t) < 1$$

Finally, the spike timing in the epoch of each trial was linearly transformed into its corresponding median of the epoch:

$$Spikes_{wrapping}(t) = \Delta\, spikes_{1:n}(t) * \eta_{Epoch}$$

Note that the algorithm enables us to preserve the number of spikes and the proportion of the inter-spike interval. The Time-warping algorithm shrinks or expands the temporal scaling by conserving the pattern and amplitude of the firing rate of the neurons.

In addition, we considered evaluating the neural activity until the fifth lick for the outcome epoch, as this parameter was found to be consistent across most trials. Including this task condition was considered important to perform a fair comparison for the subsequent neural analysis, particularly for those in which reward was delivered versus those in which it was omitted. By ensuring that both types of trials were subject to the same task constraints, we conducted a more rigorous comparison of the licking behavior on both reward outcomes.

**The neuronal responses of the NAcSh**

Following the time-warping algorithm, the firing rate of the 1507 NAcSh neurons recorded was calculated using the convolution with a Gaussian kernel (σ = 30 ms):

$$K(t|t_i\,,\ \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(t-t_i)^2}{2\sigma^2}}$$

Where $(t - t_i)$ represents a spike occurring at the time $t_i$ (i.e., timestamps of every spike per neuron). The estimate of the firing rate given by the convolution was:

$$fR(t|t_i\,,\ \sigma) = \sum_{t_i} K(t|t_i\,,\sigma)$$

Then, we normalized neuronal responses using a z-score:

$$fR_{z-score} = \frac{fR - \mu}{\sigma}$$

Where $fR$ was a time series of the firing rate of a given neuron along the trial, corresponding to 1 s before the onset of the head-in and the fifth lick delivered onto the spout in the reward port (4.16 s after the onset of the head-in). The mean and standard deviation were calculated from the firing rate within the trials.

**Outcome and Choice identification NAcSh neuronal responses**

NAcSh neurons have diverse pattern responses associated with different task variables, such as Choice (Left or Right lever) or Outcome (Delivery or Omission). The Receiver Operating Curve (ROC) analysis was implemented to quantify how different the neural firing rate for Choice (i.e., firing rate distribution during left-lever trials vs. firing rate distribution during right-lever trials) and Outcome (i.e., firing rate distribution during reward Delivery trials vs. firing rate during reward Omission trials) over time (i.e., number of consecutive bins, the bin width was systematically adjusted from 100 to 1000 ms in increments of 100 ms. Then, we chose the bin width, which reduced close to zero the proportion of neurons that were selective to both conditions (e.g., reward Delivery and Omission trials) in different time intervals. Thus, given these criteria, the optimal bin-width was 400 and 200 ms for choice and outcome, respectively.

**Permutations**

We calculated the chance threshold to assess the significance of the fraction of neurons identified using the ROC method for detecting Outcome or Choice-selective neurons. This threshold represents the number of neurons considered significant (using the same ROC criteria) purely by chance. To establish this threshold, we performed shuffling experiments on the outcome and choice variables (delivery vs. omission, left vs. right, respectively). In each iteration, we shuffled the trials, calculated the ROC, and repeated this process 1,000

times. The chance threshold for each variable was then determined based on the fraction of neurons that achieved significance through random chance.

**Lick coherent neurons**

Neuronal activity of the NAcSh exhibited spiking responses consistent with rhythmic licking behavior (within the bandwidth of 4 to 12 Hz). This coherence was identified using a multi-taper time-frequency coherence analysis.[74] The confidence intervals and the significant threshold were determined using a jack-knife method. Neuronal responses were considered lick-coherent if the lower confidence interval was above the significant threshold for only delivery, omission trials, or both (delivery & omission).[23]

**Principal Component Analysis**

The Principal Component Analysis (PCA) was implemented to understand the neural dynamics of the NAcSh population responses. Following the time-warping algorithm and the preprocessing, the neuronal responses were concatenated in an n x m matrix. Each row was the mean firing rate of a single unit, and columns were the time bins at 5 ms. The mean neuronal responses were vertically concatenated for 3 types of trials: all trials (Average), only trials where the reward was delivered (Delivery trials), and only trials where the reward was omitted (Omission trials). Linear combinations of population activity were computed, capturing the most variance in the averaged population responses. The population activity was projected onto three axes, describing the population trajectories across the task's epochs.

**Ensemble classification**

Using the same n x m, a concatenated matrix was used to classify them into clusters or, hereafter referred to as ensembles. Considering the high dimensionality of the concatenated NAcSh neuronal responses, the PCA was implemented prior to the t-distributed Stochastic Neighbor Embedding (t-SNE) using the first ten principal components embedding points taking into consideration to preserve neighborhood identity optimally.[75] Using the "Elbow curve" method, the number of clusters for the three block phases (i.e., All, Early, Late) to optimize was set to 20, and the number of ensembles found for each block condition was 14 for All, 12 for the Early phase, and 10 for the Late phase. Using the number of ensembles optimized for each block phase, neurons were grouped using a hierarchical clustering analysis, which uses the increase in the total within-cluster sum of squares of the distances

between all points in the cluster and the centroid of the cluster. The neuronal responses were grouped using a similar method as in [22].

**Methods Network inference and analysis**

Probabilistic graphical models (Markov random fields) were inferred from the electrophysiological recordings by calculating the mutual information (MI), an information theoretical measure of statistical dependency (MI is the maximum entropy or less-biased estimate of statistical dependence) using the infotheo: 'information-theoretic measures: R package version 1.2.0'.[76] After binning discretization, MI (in bits) was calculated between all neurons recorded per experiment. Once the full N x N adjacency matrix was obtained (with N the number of recorded neurons per experiment), an approach to graph sparsification was made by means of quantile thresholding. The 10% higher value tail of the mutual information distribution was determined, and all links equal to or above this threshold were retained while discarded elsewhere. The set of links and nodes retained form a probabilistic network for each condition.

Networks were visualized, and their resulting topological structure was analyzed using Cytoscape[77] to determine local and global centrality measures. Connectivity degree distribution, local and global clustering coefficients, average shortest path length, and betweenness centrality distributions were also determined using the Network Analyzer package in Cytoscape.

**Data Availability Statement**

The datasets presented in this article are available on request from the corresponding author. Requests to access the datasets should be directed to RG,

**Ethics Statement**

CINVESTAV Animal Care and Use Committee reviewed and approved the animal study.

Author Contributions

BA and RG designed the research. BA performed the research. BA, EHL, and RG analyzed the data, BA, EHL, and RG wrote the paper. All authors reviewed and approved the manuscript for publication.

**Funding**

This project was supported in part by Productos Medix 3247, Cátedra Marcos Moshinky, CONHACyT CF-2023-G-518 to RG.

**Conflict of Interest**

The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

**Acknowledgments**

Benjamin Arroyo is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received a fellowship from CONHACyT, México, and data in this work is part of his doctoral dissertation in the Posgrado en Ciencias Biomédicas of the Universidad Nacional Autónoma de México. We thank Mario Gil Moreno for building homemade electrodes. We also thank Fabiola Hernandez Olvera for her invaluable animal care.

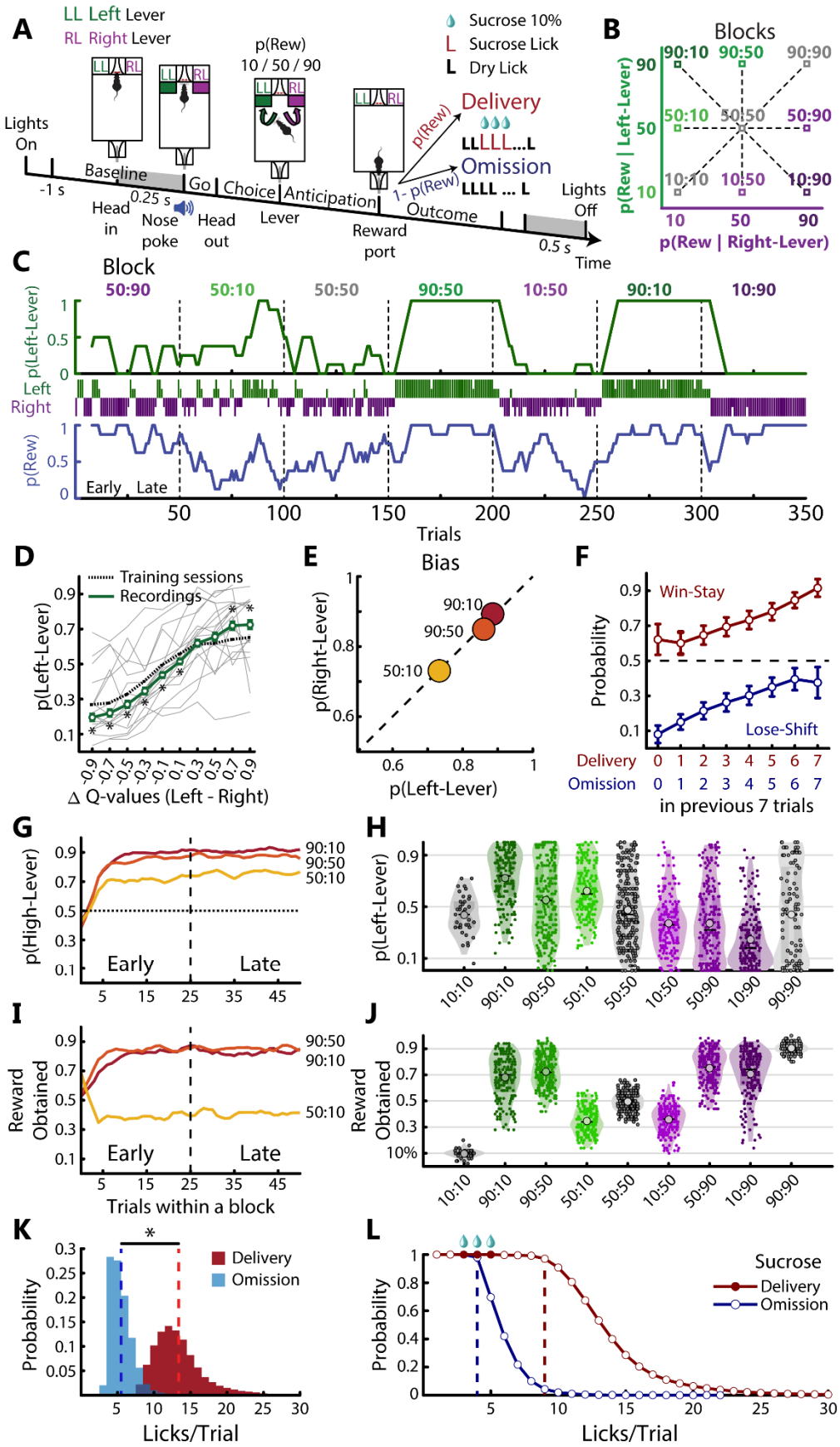

A
LL Left Lever
RL Right Lever
p(Rew)
10 / 50 / 90
Sucrose 10%
L Sucrose Lick
L Dry Lick
Delivery
LLLLL...L
Omission
LLLL ... L
p(Rew)
1- p(Rew)
Lights On
-1 s
Baseline
0.25 s
Go
Choice
Anticipation
Outcome
Lights Off
0.5 s
Time
Head in
Nose poke
Head out
Lever
Reward port
B
Blocks
90:10
90:50
90:90
50:10
50:50
50:90
10:10
10:50
10:90
p(Rew | Left-Lever)
p(Rew | Right-Lever)
C
Block
50:90
50:10
50:50
90:50
10:50
90:10
10:90
p(Left-Lever)
Left
Right
p(Rew)
Early
Late
Trials
D
Training sessions
Recordings
p(Left-Lever)
Δ Q-values (Left - Right)
E
Bias
90:10
90:50
50:10
p(Right-Lever)
p(Left-Lever)
F
Win-Stay
Lose-Shift
Probability
Delivery 0 1 2 3 4 5 6 7
Omission 0 1 2 3 4 5 6 7
in previous 7 trials
G
p(High-Lever)
90:10
90:50
50:10
Early
Late
H
p(Left-Lever)
10:10
90:10
90:50
50:10
50:50
10:50
50:90
10:90
90:90
I
Reward Obtained
90:50
90:10
50:10
Early
Late
Trials within a block
J
Reward Obtained
10%
K
Delivery
Omission
Probability
Licks/Trial
L
Sucrose
Delivery
Omission
Probability
Licks/Trial

**Figure 1. Design and performance of a self-guided probabilistic choice task.**

**(A)** A single trial structure: A trial began with the lights on, and the box illumination instructed rats to nose poke the central port. An auditory tone (880 Hz and 80 dB) was delivered if the rat remained in that position for 0.25 s. The Baseline epoch comprised -1 to 0.25 s from head entry. At the onset of the Go signal, two levers were extended, and the rat headed out from the central port (Go epoch). In the Choice epoch, rats were required to choose and press one of the two levers (left-green or right-purple color). Both levers were retracted after each choice, and the Anticipation epoch began. In this epoch, rats needed to approach the reward port on the opposite side of the box. During the Outcome epoch, when a trial was rewarded, after two dry licks, a drop of 10% sucrose was delivered in the following three licks, one drop per lick (Delivery, see the three red L's and the three drops). Otherwise, the reward was omitted (Omission trial). The trial was finished 500 ms after rats refrained from licking and the light was turned off. From block to block, rats had to infer the reward probability associated with each lever to self-guided choices and update their beliefs every 50 trials. The transition between blocks was not cued.

**(B)** Diagram of all reward probabilities tested. For instance, in block 90:10, 90% indicates the reward probability for the left and 10% for the right (see squares). Blocks with higher reward probability for the left lever are represented in green, purple for the right lever, and gray for blocks where the reward probabilities are equal for both levers.

**(C)** Choice behavior within a representative session. The top plot is the block sequence throughout a session. The green line depicts choices on the left lever (smoothed over 8 trials window). Green and purple vertical ticks denote left and right lever choices in the middle, respectively. Long ticks indicate reward delivery, whereas short ticks are reward omissions. Bottom, the blue line is the reward rate. We considered the early phase the first half of every block (trials from 1 to 25) and the late phase the second half (trials from 26 to 50). The dashed vertical lines indicate block transitions.

**(D)** Left-choice probability as a function of action-value difference. Action values were computed using the Q-learning model. Q-values near 1 indicate that the rat is more likely to choose that lever. Hence, the left lever is more likely to be pressed if the difference between Q-values (Q-value Left - Q-value Right) is positive. As predicted by the model, rats choose left when the Q-value of the left lever is higher than the right lever. Note that the probability of choosing the left lever further improved after the electrodes were implanted in the NAcSh

(*p<0.05 recording vs. training unpaired t-test, solid-green and dashed-black sigmoid, respectively). Individual rats are shown in gray lines.

**(E)** Comparison between leftward and rightward blocks. No biases were found in the asymmetric blocks.

**(F)** Given that no sensory cue could predict the outcome, it forces animals to integrate multiple previous trials to guide choice and promote a win-stay and lose-shift strategy [30]. In a win-stay strategy, the likelihood of choosing the same lever increased as the rat received more rewards in the past. In contrast, the likelihood of choosing a different option increases after reward omission in the lose-shift strategy.

**(G)** Probability of choosing the lever with higher reward probability as a function of trials within the block. The horizontal dotted line indicates chance level (0.5), and the vertical dashed line is the transition from the Early to Late phase within a block.

**(H)** Probability of choosing the left lever throughout blocks. Each small dot depicts a single session block; shaded areas are the kernel density function. Horizontal lines and big dots indicate the median and the mean, respectively. The left lever is more likely to be chosen in blocks with more reward probability in that lever.

**(I)** A reward was obtained for the lever with high probability across trials in the block.

**(J)** The overall rewards obtained per block.

**(K)** Distribution of licks given in the reward port during reward delivery and omission (number of licks for delivery trials: 13.39 ± 0.02, and omission trials: 5.5 ± 0.01, vertical dashed lines, mean ± SEM).

**(L)** Distribution of lick probability given in the reward port during sucrose delivery and omission. Rats rapidly detect reward omissions in about one more lick, so we forced them to give at least 4 licks to finish an omission trial. Under these conditions, rats learned not to lick the empty sipper at the 4$^{th}$ empty lick (see blue dashed line). In contrast, in sucrose delivery trials, rats engaged in rhythmic licking for at least 9 licks (red dashed line) before gradually decreasing the probability of being given more licks.

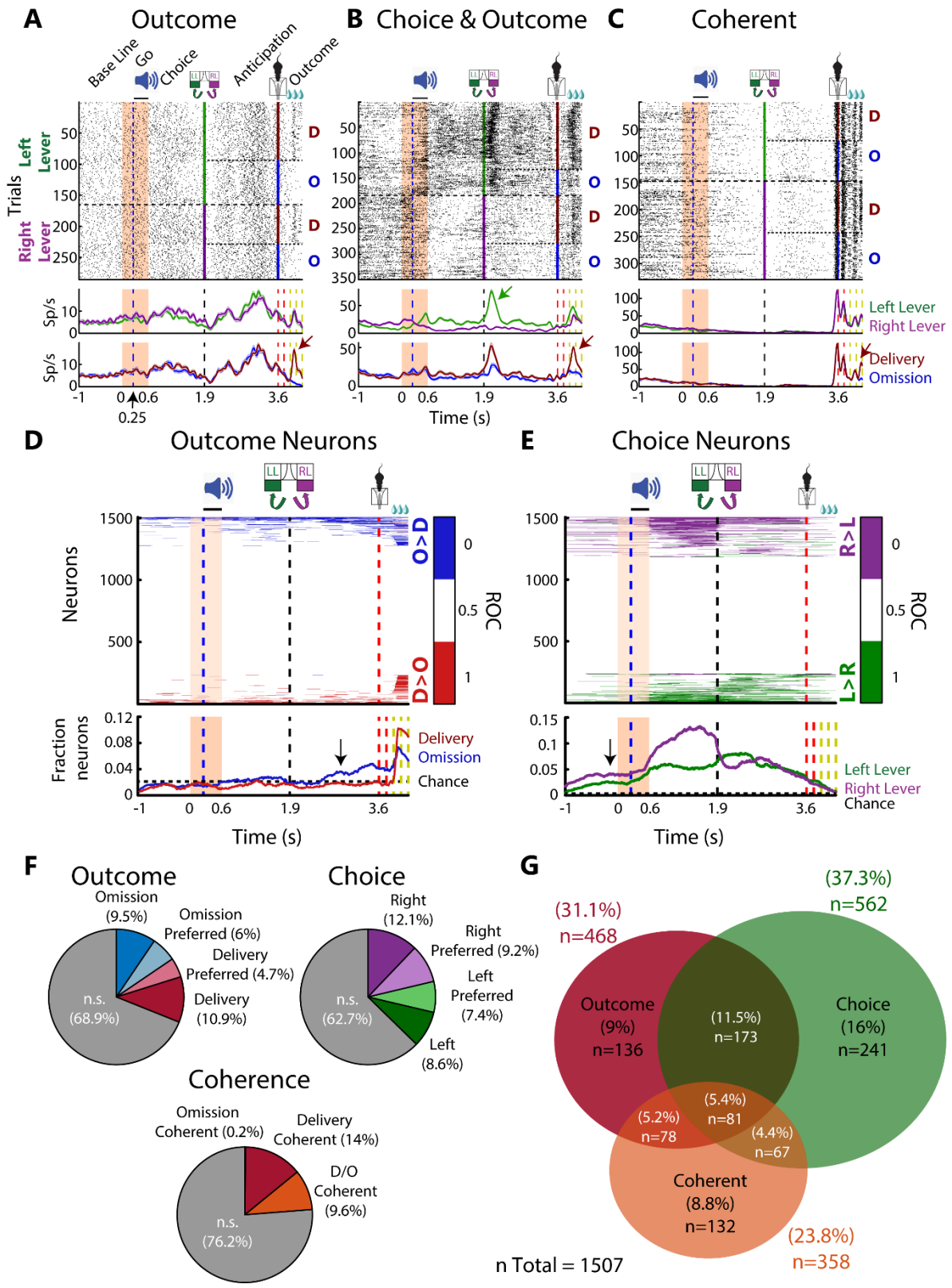



**Figure 2. Extracellular single-unit recordings identified NAcSh responses that tracked the choice, reward outcome, and rhythmic licking.**

**(A)** Single neuron responses can vary across trials hidden some spike patterns; Time warping of the neuronal responses allows us to visualize the precise temporal structure across epochs (**Figure 2 supplementary 1**)[33]. The left raster depicts the time-warped response of an Outcome neuron that fired more to the delivery than the omission of sucrose (see red arrow). Raster plots were aligned to head entry (Time = 0 s). The blue dashed line marked the "Go" tone (Time = 0.25 s), and the head out is shown at time = 0.6 s. Lever presses are displayed at 1.9 s (green for the left and purple for the right lever). All blocks were pooled together, and trials were sorted by left (green) or right lever (purple vertical line) and then by reward delivery (D; red ticks) or omission (O; blue marks). The letters D and O stand for delivery and omission. The 1st lick delivered in the reward port is shown at 3.6 s: Red dashed lines are empty licks, yellow are sucrose only for delivery trials (3 drops / one per lick), whereas in omission are empty. The middle panel indicates the PSTH for the left lever (green) and the right lever trials (purple). The bottom panel is the PSTH responses sorted by delivery (red) or omission (blue) trials. For details of the preprocessing time warping activity, see **Figure 2 supplementary 2**.

**(B)** A representative NAcSh neuron selective for the left lever (see green arrow) and sucrose delivery (red arrow). Such responses were classified as choice & outcome neurons.

**(C)** Phase-locked firing of a representative neuron entrained by licking a spout in the reward port. This lick-coherent neuron exhibited oscillatory activity that decreased amplitude as the licks were given in the reward port (see dashed and yellow lines). This type of response was most likely related to the oromotor aspects of licking rather than gustatory information. Same convention as panel (**A**).

**(D)** Outcome neurons identified by a receiving operating characteristic (ROC) curve. In blue indicates bins when the neuron exhibited a higher firing rate to omission than delivery trials (O>D: omission greater than delivery). In comparison, red denotes significant bins, where the firing rate was significantly higher for reward delivery than omission (D>O: delivery greater than omission).

**(E)** ROC analysis for Choice neurons: in purple, neurons were more selective for the right than the left lever (R>L), and green preferred the left lever (L>R).

**(F)** The recorded NAcSh neurons were examined for significant modulation by Outcome, Choice, and Coherence. Outcome neurons were classified as Delivery (10.9%), firing consistently across all task epochs for delivery trials (D>O), whereas Omission neurons

(9.5%) fired at a higher rate during omission than delivery trials across all task epochs. Some neurons encoded both delivery and omission at different task epochs, classified as “preferred” when more bins were significantly higher for Delivery (Delivery preferred 4.7%) or Omission (Omission preferred 6%). The pie chart at the right depicts similar categorizations for Choice neurons. For Coherent neurons, Omission Coherent (0.2%) or Delivery Coherent (14%) were further subdivided. D/O Coherent neurons (9.6%) exhibited lick coherence in both delivery/omission trials.

**(G)** Veen’s diagram illustrates the logical relations between Outcome, Choice, and Coherent-related neuron sets. Among the 1507 neurons recorded, 31.1% (n = 468) showed outcome-related activity, 37.3% (n = 562) exhibited choice-related activity, and 23.8% (n = 358) displayed licking coherent. As these categories are not mutually exclusive, a significant portion of NAcSh neurons simultaneously responded to two or all three variables.

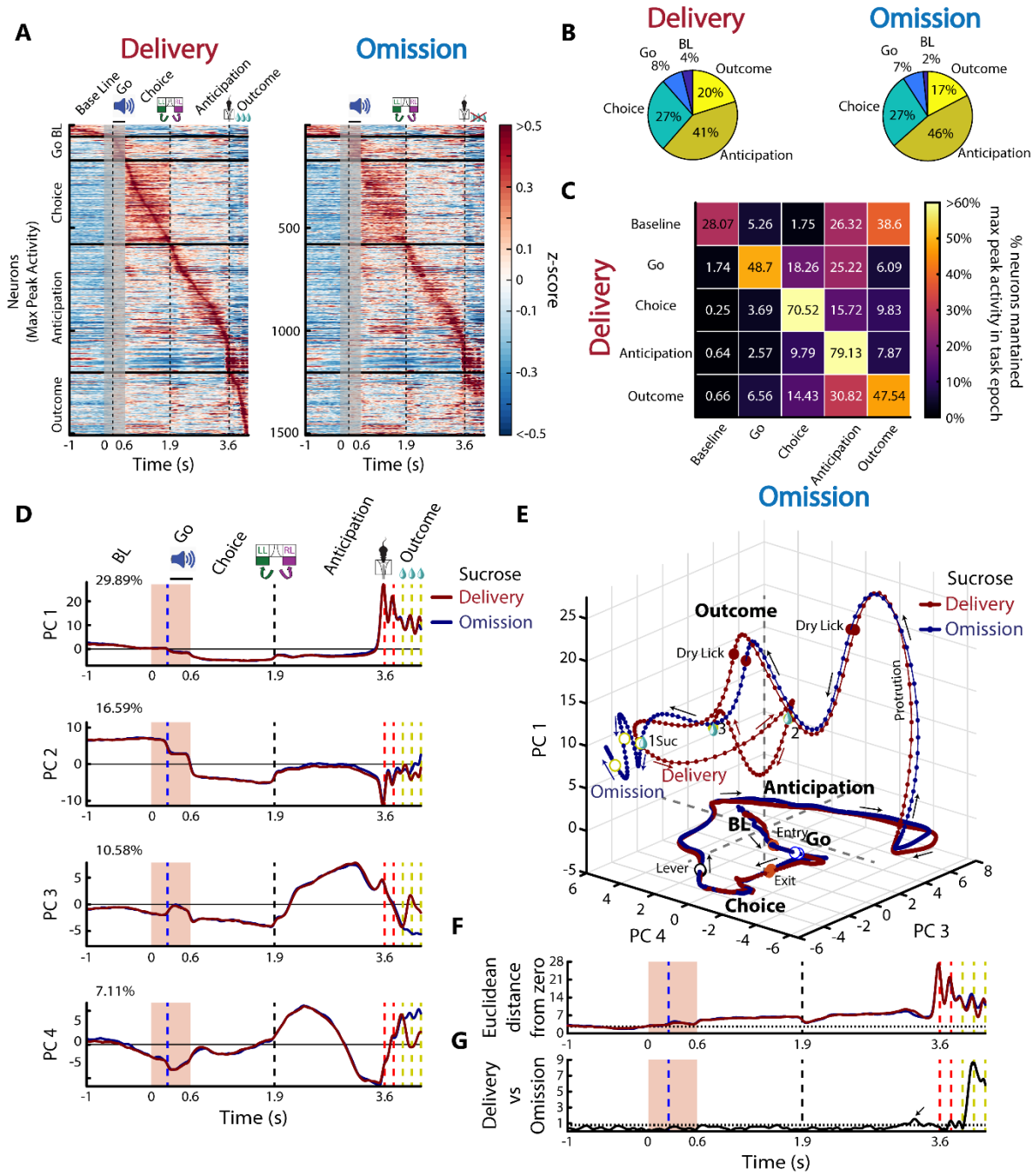



**Figure 3. Peak activity of NAcSh neurons sequentially represents all task epochs and PCA trajectories discriminated between reward outcomes.**

**(A)** In the left panel, individual neurons were sorted by their maximum peak in trials where the reward was delivered. The same ranking index was used to sort the neuronal responses, but now, during reward omission (right panel). All responses were aligned (time = 0 s) to the

head entry (onset of gray shaded area). From left to right, the first dashed line represents the Go signal, the second the lever press, and the third the first empty lick in the reward port.

**(B)** The percentage of NAcSh neurons exhibiting peak firing rate differed across task epochs. These neurons peaked mainly during Anticipation and Choice, followed by Outcome. Baseline and Go epochs had fewer neurons with peak activity.

**(C)** Neuronal peak persistence across reward outcomes: This analysis quantifies the proportion of neurons that maintain peak activity during the same epoch when the reward was delivered or omitted. Interestingly, the Choice and Anticipation epochs recruited a higher percentage of neurons whose peak activity persisted across both reward outcomes. Conversely, neurons in the Baseline, Go, and Outcome epochs primarily shifted their peak activity to the Anticipation epoch.

**(D)** Principal Component Analysis (PCA) of neuronal population dynamics across task epochs during reward Delivery (red) and Omission (blue). Note that PC3 and PC4 components distinguished between reward outcomes. The first four PCs explain a total variance of 69.31%. The percentage of the variance explained by each principal component is also shown.

**(E)** The PCA trajectories for Delivery and Omission trials (red and blue dotted lines, respectively), in which population activity was described as a trajectory through spatiotemporal activity patterns.[35] The arrows indicate the direction that trajectories followed, and they began from the Baseline (BL), Go, Choice, Anticipation, and Outcome epoch. **Video Fig. 3 Supplementary 1** details PCA trajectories as a function of neuronal ensembles. More of this in **Figure 4**.

**(F)** Euclidean distance for both time series in Delivery and Omission trials (red and blue lines, respectively). Initially, during the baseline, both trajectories were near the subspace's origin. As the trial progresses, they gradually move away, with the most significant variance occurring during licking in the outcome epoch. The horizontal dotted line represents the mean distance value during the baseline epoch.

**(G)** The separation of reward vs. oromotor effects is illustrated using the NAcSh PCA trajectories. The solid black line represents the Euclidean distance between outcome trajectories. The primary effect occurs precisely during the delivery or omission of sucrose, independent of oromotor differences. An arrow marks the effect during the anticipation

epoch. The horizontal dotted line represents the maximum distance value during the baseline.

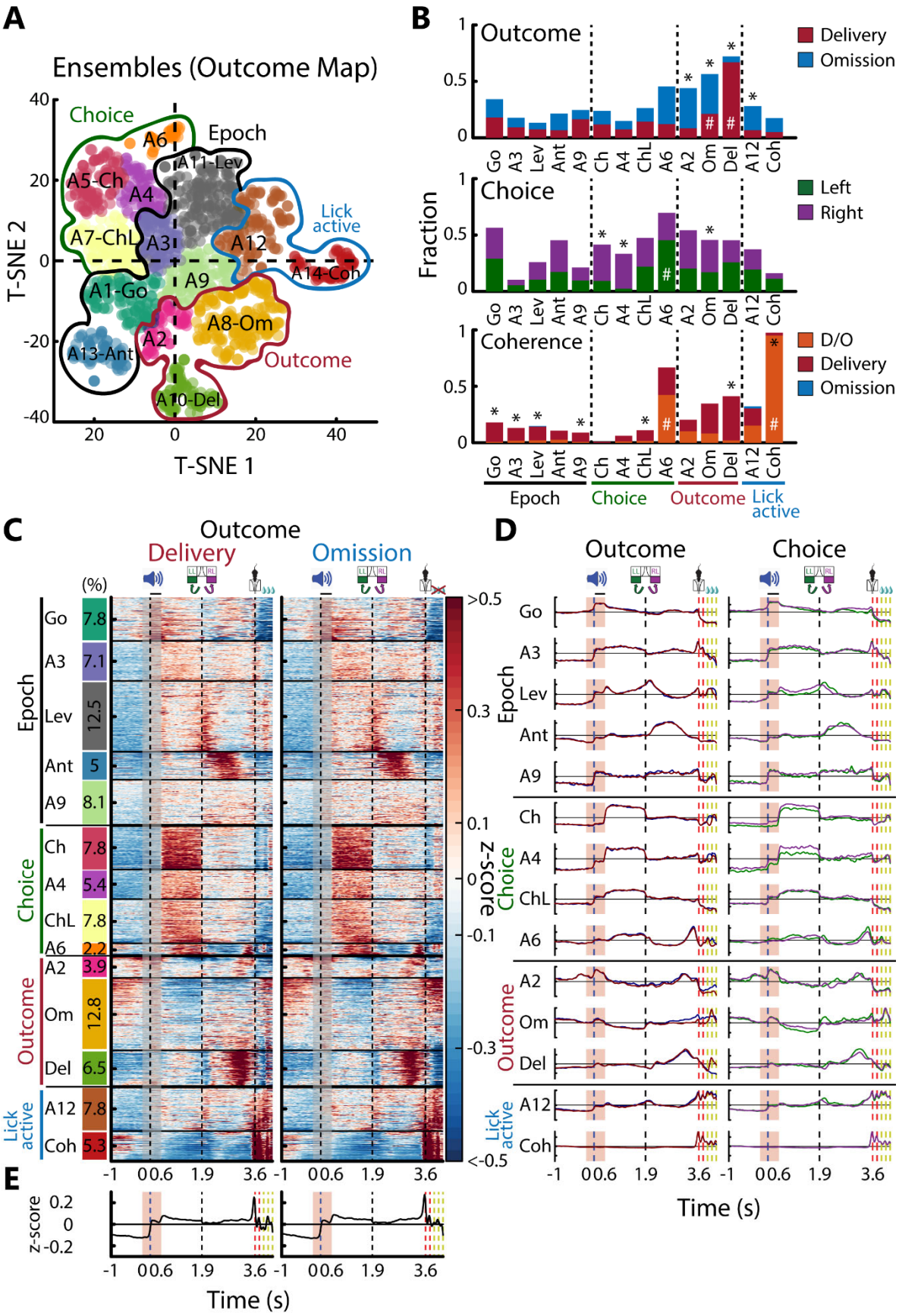

A
Ensembles (Outcome Map)
Choice
Epoch
Lick active
Outcome
A6
A11-Lev
A5-Ch
A4
A7-ChL
A3
A12
A14-Coh
A9
A1-Go
A8-Om
A2
A13-Ant
A10-Del
T-SNE 2
T-SNE 1
B
Outcome
Delivery
Omission
Choice
Left
Right
Coherence
D/O
Fraction
Go A3 Lev Ant A9 Ch A4 ChL A6 A2 Om Del A12 Coh
Epoch
Choice
Outcome
Lick active
C
Outcome
Delivery
Omission
(%)
Epoch
Go 7.8
A3 7.1
Lev 12.5
Ant 5
A9 8.1
Choice
Ch 7.8
A4 5.4
ChL 7.8
A6 2.2
Outcome
A2 3.9
Om 12.8
Del 6.5
Lick active
A12 7.8
Coh 5.3
>0.5
0.3
0.1
0
-0.1
-0.3
<-0.5
z-score
-1 0 0.6 1.9 3.6
D
Outcome
Choice
Epoch
Go
A3
Lev
Ant
A9
Choice
Ch
A4
ChL
A6
Outcome
A2
Om
Del
Lick active
A12
Coh
Time (s)
E
z-score
0.2
0
-0.2
Time (s)

**Figure 4. The diverse activity patterns of NAcSh responses were grouped into 14 neuronal ensembles, each specializing in processing specific task- and reward-related information.**

**(A)** The Outcome ensemble map was generated using t-distributed stochastic neighbor embedding (t-SNE) of time-warped firing rates (see Methods). T-SNE, a statistical method for visualizing high-dimensional data, assigns each data point (neuron) a location on a two-dimensional map (referred to as the NAcSh Outcome ensemble map). Each dot represents a single neuron, with color indicating its assigned ensemble.

**(B)** Bar plot illustrating the fraction of neurons in each ensemble carrying information about the Outcome (top), Choice (middle), and Coherence (bottom) variables. Ensembles are sorted based on neural activity patterns, forming different Meta-ensembles: Epoch (black), Choice (green), Outcome (red), and Lick active (blue). The hash symbol denotes ensembles with a larger fraction of neurons selective to a variable compared, while the asterisk symbol indicates statistically significant differences between conditions (e.g., delivery vs. omission, left vs. right). **Outcome**: Ensembles A8-Om and A10-Del are enriched with neurons encoding outcome variables. Stacked bars show the fraction of neurons selective for reward delivery or omission (red and blue bars, respectively). **Choice:** The choice variable distribution is more uniform across ensembles, with green and purple bars representing the fraction of neurons selective to left and right choices. **Coherence**: Neurons in ensembles A6 and A14-Coh exhibit activity significantly correlated with rhythmic licking behavior. Red and blue bars correspond to coherence during reward delivery and omission, respectively, while orange indicates coherence.

**(C)** Colored-coded heatmaps for neurons assigned to fourteen ensembles. Each row represents the z-scored mean activity from a single neuron during reward delivery (left) and omission trials (right). Neuronal ensembles are sorted as in **B**, and the vertical bar color on the left indicates the percentage of neurons for each neuronal ensemble. All responses were aligned to the head entry (onset of gray shaded area). From left to right, the first vertical dashed line is the Go signal, the second the lever press, and the third is the first lick delivered in the reward port. Solid, thick horizontal lines separate individual ensembles and meta-ensembles, respectively.

**(D)** Average ensemble activity with distinct neuronal activity patterns identified by t-SNE. For Outcome (left), ensemble activity is segregated by reward delivery and omission (red and

blue lines), while for Choice (right panel), activity is sorted by the left and right choices (green and purple lines). Neuronal ensembles include Go (A1-Go), A3, Lev (A11-Lev), Ant (A13-Ant), A9, Choice (A5-Ch), A4, Choice-Late (A7-ChL), A6, A2, Omission (A8-Om), Delivery (A10-Del), A12, and Coherent (A14-Coh).

**(E)** The black line represents the average population activity in the NAcSh for reward delivery (left) and omission trials (right).

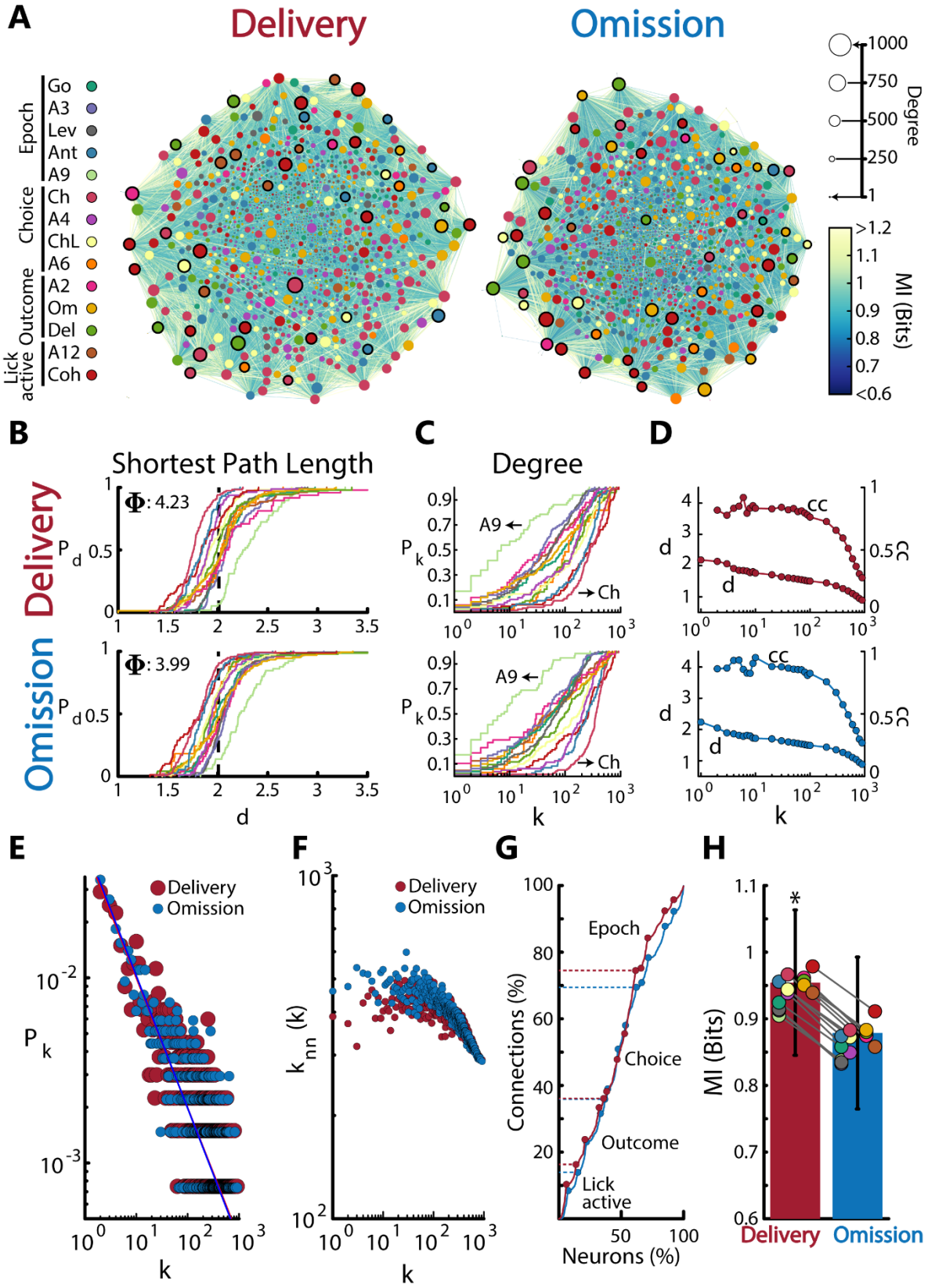

A
Delivery
Omission
Epoch
Go
A3
Lev
Ant
A9
Choice
Ch
A4
ChL
A6
Outcome
A2
Om
Del
Lick active
A12
Coh
1000
750
500
250
1
Degree
>1.2
1.1
1
0.9
0.8
0.7
<0.6
MI (Bits)
B
Shortest Path Length
Delivery
Omission
Φ: 4.23
Φ: 3.99
P d
d
C
Degree
A9
Ch
P k
k
D
cc
d
E
Delivery
Omission
P k
k
F
k nn (k)
G
Epoch
Choice
Outcome
Lick active
Connections (%)
Neurons (%)
H
MI (Bits)
Delivery
Omission

**Figure 5. The NAcSh activity follows a small-world architecture with scale-free-like properties, and reward delivery enhances the mutual information shared across the network.**

**(A)** The NAcSh network during reward delivery and omission. The network is illustrated by an organic layout (computed with Cytoscape), in which each neuron is represented by a node (circle) and a 'functional' connection by a line between a pair of nodes. Nodes are color-coded based on their respective ensemble color. The size of the node depicts the connectivity degree. The color of the connections indicates the Mutual information shared among nodes (see **Figure 5 Supplementary 1.** for the complete adjacency MI matrix). Nodes with a thick black border represent nodes with high betweenness centrality, defined as how much a neuron is within the shortest paths between any pair of nodes in the network.

**(B)** Cumulative density function of the average shortest path length (SPL) in the NAcSh network. Solid lines are colored-coded and represent their respective ensemble's distance cumulative density function during reward delivery (top panel) and omission (bottom panel). A vertical black dashed line indicates the mean distance in the NAcSh network (for reward delivery and omission). The network diameter was denoted by Φ, which reflects the largest shortest path length, hence the network size.

**(C)** Cumulative density function of degree connectivity of covariances in the NAcSh network. Given the high variability in the nodes' degree values, the degree was plotted on a log scale. Solid lines are colored-coded and represent their neuronal ensemble degree cumulative density function. Note that some neuronal ensembles had a higher probability of containing well-connected nodes or hubs (see Ch) than others (A9).

**(D)** The covariance network in the NAcSh follows a small-world architecture. The distance (or SPL, denoted by "d" and the left Y-axis) and Closeness Centrality (denoted by "cc" and the right Y-axis) as a function of degree connectivity of the nodes in the network during reward delivery (top panel) and omission (bottom panel). As the distance between the nodes in the network suddenly decreases, the closeness centrality remains unaffected as the node's degree of connectivity increases. These measurements reflect the neuronal interconnectivity in the NAcSh at a global (SPL) and local (cc) level, facilitating the flow of information.

**(E)** A heavy-tailed distribution was found in the NAcSh network. Circles in red and blue depict the degree distribution for reward delivery and omission. A power-law function was

fitted (Delivery γ=0.69, Omission γ=0.70). The degree distribution of the covariances in NAcSh activity followed a long tail distribution with a few highly connected hubs (see **SUPPLEMENTAL INFORMATION**).

**(F)** A neutral topology was found in the NAcSh network. The degree correlation for the NAcSh network during the reward delivery and omission is depicted by circles in red and blue, respectively. A flat distribution of the connections was found throughout the nodes based on the degree of connectivity, suggesting that hubs do not tend to connect or avoid between them. The $K_{nn}(k)$ is a degree correlation function to help us capture the absence of connection patterns between nodes as a function of the node's degree connectivity (denoted by k).

**(G)** Similar connection distribution across ensembles in the NAcSh network. Percentage of connections across neuronal ensembles in the NAcSh network during reward delivery and omission (red and blue lines, respectively). Horizontal dashed lines represent the fraction of connections throughout the meta-ensembles, while dots indicate the number of connections across neuronal ensembles.

**(H)** A stronger Mutual Information was observed in the NAcSh network during sucrose consumption. Red and blue bars depicted the Mutual Information in the NAcSh network during trials when the reward was delivered or omitted, respectively. Solid black lines indicate the SEM. Circles are color-coded based on their neuronal ensemble and indicate the average Mutual Information shared in the network during reward delivery (left) and omission (right).

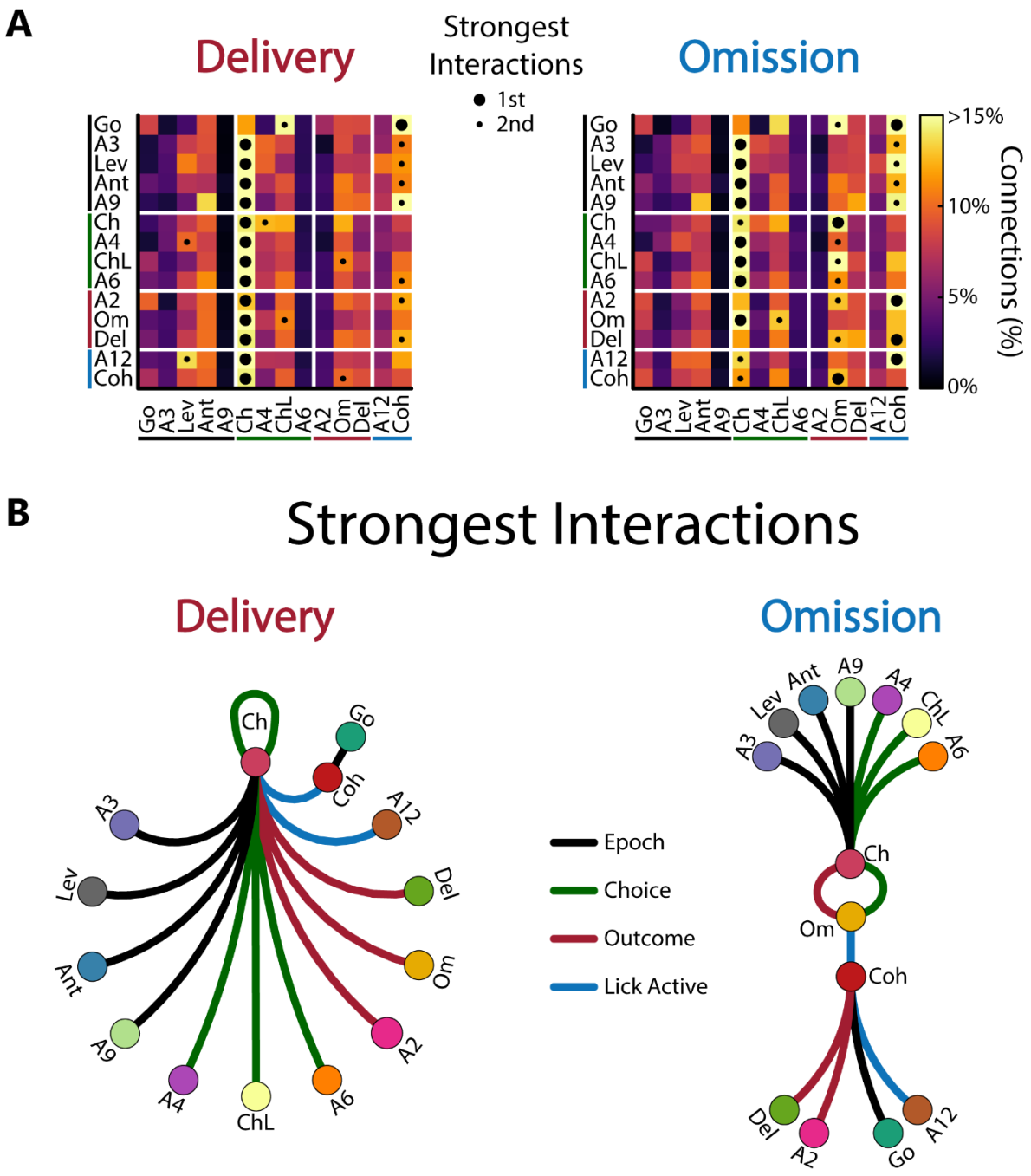



**Figure 6. The interaction backbone between NAcSh ensembles during reward delivery and omission**

**(A)** The heatmaps display the connectivity matrix between neuronal ensembles during reward delivery (left) and omission (right). Each row and column represent the percentage of total connections accounted pair-wise between neuronal ensembles. Black dots depict pairs of neuronal ensembles, with a big dot indicating the 1st strongest interaction and a

small dot indicating the 2nd strongest interaction. Both axes are sorted based on meta-ensembles Epoch (black line), Choice (green line), Outcome (red line), and Lick Active (blue line). The heatmap is divided as the meta-ensembles by horizontal and vertical solid white lines.

**(B)** Graphical representation of the 1st strongest interaction in the NAcSh network during reward delivery (left panel) and omission (right panel). Each circle represents a neuronal ensemble. The circle size represents the Mutual Information shared for each neuronal ensemble. The strongest interactions are indicated by thick lines and colored with their respective meta-ensemble color. Note that neuronal ensemble interactions are different for each outcome. During delivery trials, the Ch ensemble interacts with all the other neuronal ensembles, including itself, whereas in omission trials, The Om ensembles interact with Ch and Coh ensembles, and Ch strongly interacts with epoch meta-ensemble and other ensembles in choice meta-ensemble. Moreover, Coh ensemble interacts with A12, Go, Del, and A2.

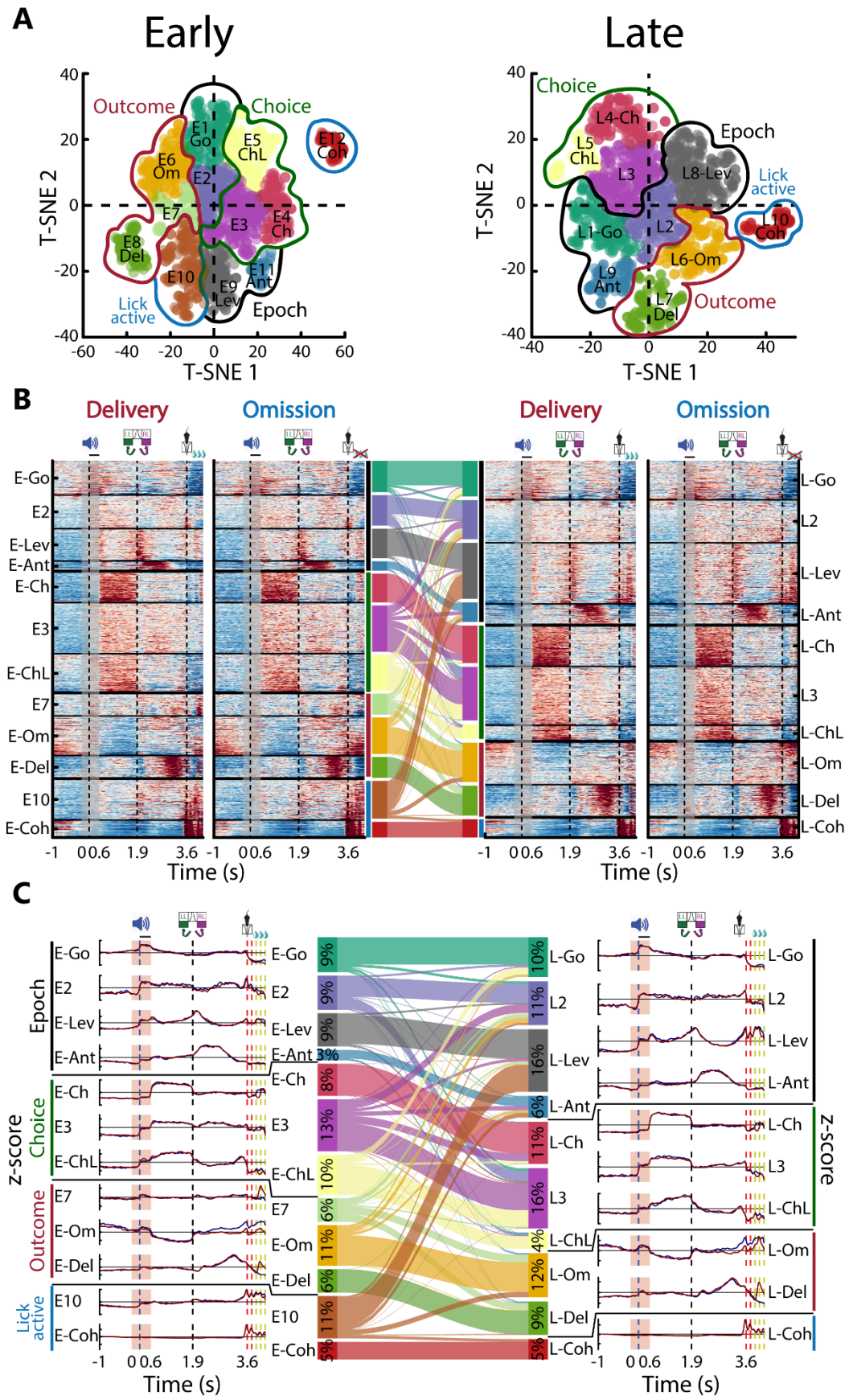

A
Early
Late
Outcome
Choice
Epoch
Lick active
E1 Go
E2
E3
E4 Ch
E5 ChL
E6 Om
E7
E8 Del
E9 Lev
E10
E11 Ant
E12 Coh
L1-Go
L2
L3
L4-Ch
L5 ChL
L6-Om
L7 Del
L8-Lev
L9 Ant
L10 Coh
T-SNE 1
T-SNE 2
B
Delivery
Omission
E-Go
E2
E-Lev
E-Ant
E-Ch
E3
E-ChL
E7
E-Om
E-Del
E10
E-Coh
L-Go
L2
L-Lev
L-Ant
L-Ch
L3
L-ChL
L-Om
L-Del
L-Coh
Time (s)
C
z-score
Epoch
Choice
Outcome
Lick active
9%
9%
9%
3%
8%
13%
10%
6%
11%
6%
11%
5%
10%
11%
16%
6%
11%
16%
4%
12%
9%
5%

**Figure 7. Neuronal ensembles are dynamic, with neurons being removed from and added to other ensembles as reinforcement learning progresses.**

**(A)** t-SNE maps for neural activity in the Early (first 25 trials) and Late (last 25 trials) task phases. Ensemble dynamics in the NAcSh exhibit temporal shifts across task phases. t-SNE and hierarchical clustering revealed a greater number of ensembles (n=12) during the Early phase (first 25 trials) compared to the Late phase (n=10 ensembles; last 25 trials), suggesting a dynamic reorganization of neuronal activity patterns as the task progresses. The ensembles found in the Early phase were named with the initial letter "E," whereas those found in the Late phase were tagged with an "L."

**(B)** In the left panel, the heatmap illustrates the z-scored activity of each neuron sorted as a function of the 12 ensembles identified during the early phase. Same conventions as in **Figure 4C**. In the middle panel, the Sankey diagram draws the transition of individual neurons in each ensemble from the Early to Late phase. Right panel, the heatmap displays the z-scored activity pattern of the now 10 ensembles identified using the neural activity of the late phase. Ensembles in both task phases are sorted based on the meta-ensembles they belong to.

**(C)** Left panel, PSTH of average ensemble activity across the 12 neuronal ensembles identified in the Early phase. The red and blue lines represent the reward delivery and omission activity. Same conventions as **Figure 4D**. In the middle panel, the Sankey diagram draws on how neurons drop in and out between neuronal ensembles from the Early to Late phase, and the percentage of neurons for each ensemble is shown on both sides of the diagram. In the right panel, the PSTH of the average activity pattern through the 10 neuronal ensembles is now found in the Late phase. Neuronal ensembles in both phases are sorted, as in panel **B**.

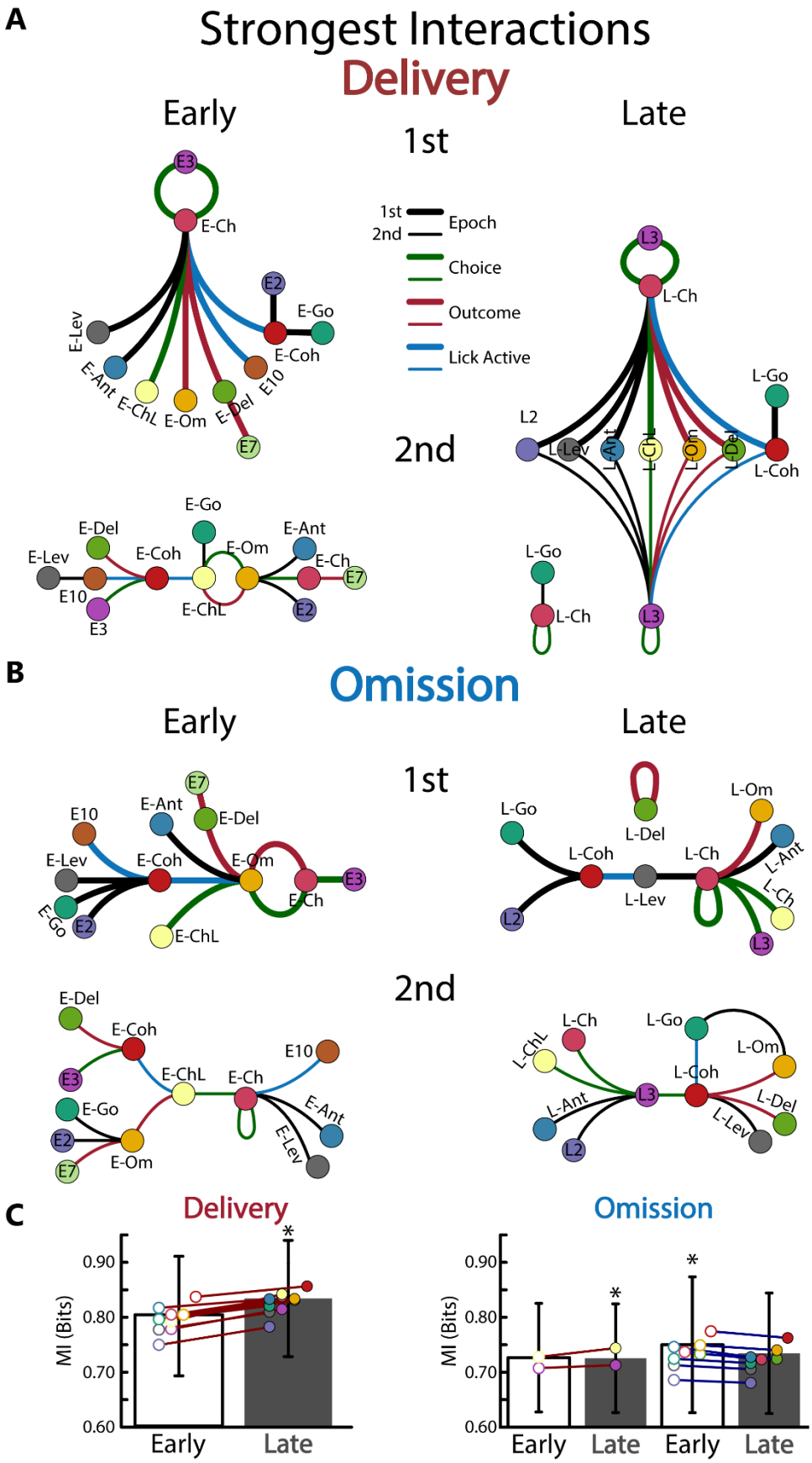

A
Strongest Interactions
Delivery
Early
Late
1st
2nd
1st
2nd
Epoch
Choice
Outcome
Lick Active
E3
E-Ch
E2
E-Go
E-Lev
E-Coh
E-Ant
E-ChL
E-Om
E-Del
E10
E7
L3
L-Ch
L-Go
L2
L-Lev
L-Ant
L-ChL
L-Om
L-Del
L-Coh
B
Omission
Early
Late
1st
2nd
C
Delivery
Omission
MI (Bits)
0.90
0.80
0.70
0.60
Early
Late

**Figure 8. Reinforcement learning progressively reorganizes the strength of connectivity during the reward delivery and omission**

**(A)** The NAcSh network reshapes its connectivity during reward delivery. Circles are colored by the corresponding neuronal ensemble. The circle size depicts the mutual information shared by the neuronal ensemble. Thick and thin lines depict the strongest first (top) and second (bottom) interactions, respectively. The color line of the interactions indicates the Meta-Ensemble to which it belongs.

**(B)** Functional connectivity reorganization in the Omission network. Same convention as in panel (**A**).

**(C)** Reward delivery and omission underwent opposite strength connectivity while learning a reinforcement learning task. The left panel exhibited an increase in mutual information in the NAcSh network during trials when the sucrose was delivered (MI Reward delivery, Unpaired t-test: Early vs. Late, t-value $_{(225,596)}$ = -70.92, $p<0.0001$). Right panel exhibited two separated comparisons: the first shows two neuronal ensembles that increased the mutual information (neuronal ensemble 3 and ChL, MI Reward omission, Unpaired t-test: Early vs. Late, t-value $_{(81,133)}$ = -9.37, $p<0.0001$). While the second comparison exhibited the remaining neuronal ensembles that collectively decreased the mutual information during the reward-related learning (MI Reward omission, Unpaired t-test: Early vs. Late, t-value $_{(134,102)}$ = 28.27, $p<0.0001$)

**Supplementary Figures**

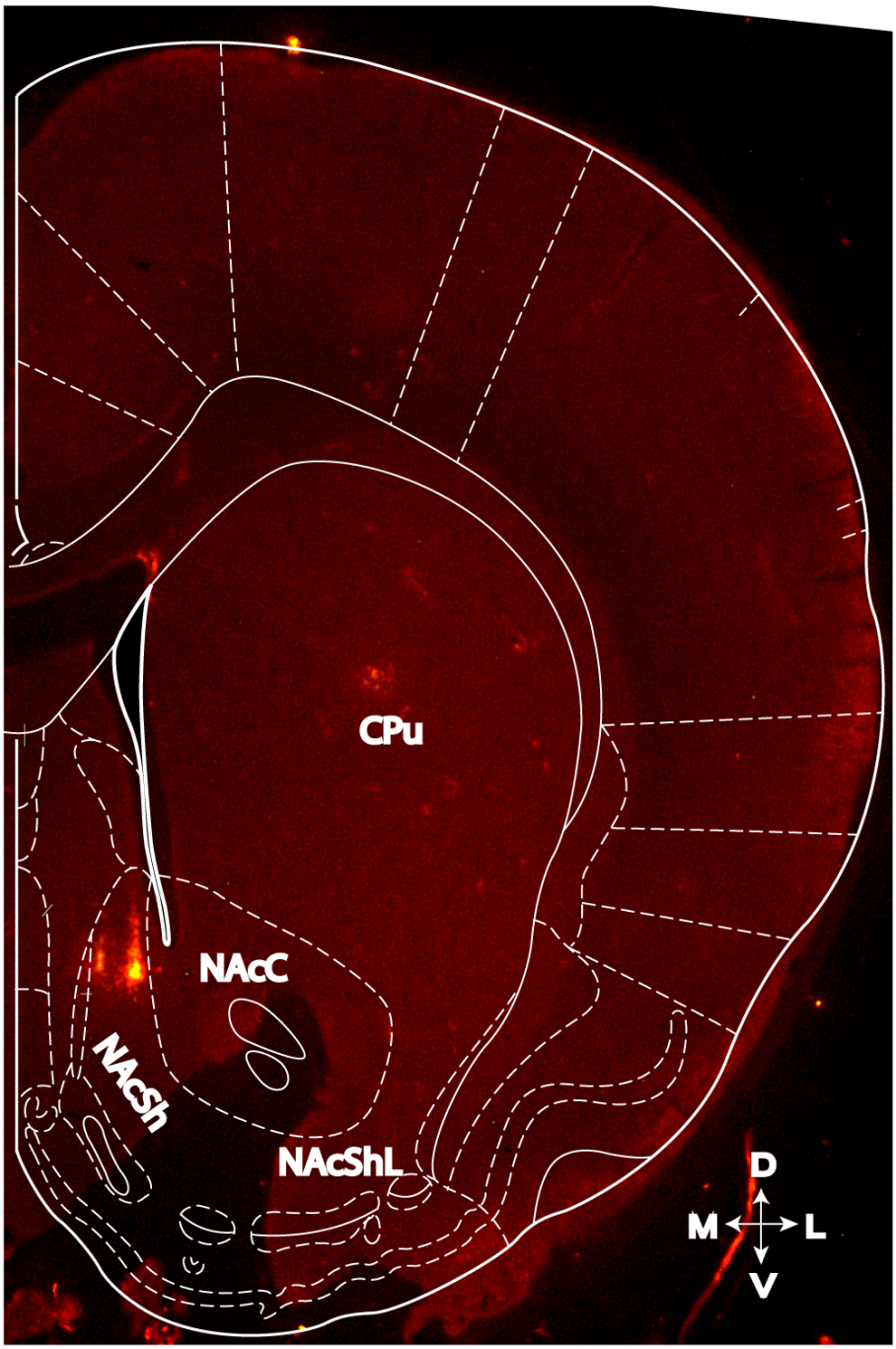



**Figure 2 Supplementary 1. A representative track of recording sites.**

A coronal mouse brain section was viewed in an epifluorescence microscope in which the electrode array was stained by a fluorescent dye (cell tracker CM-DiI dye, red excitation/emission spectra: 553/570 nm; ThermoFisher Scientific, MA, USA). The electrode array was found in the anterior and medio-dorsal parts of the Nucleus Accumbens Shell

(NAcSh). NAcC = nucleus accumbens core and NAcShL = nucleus accumbens lateral. CPu, = Caudate putamen

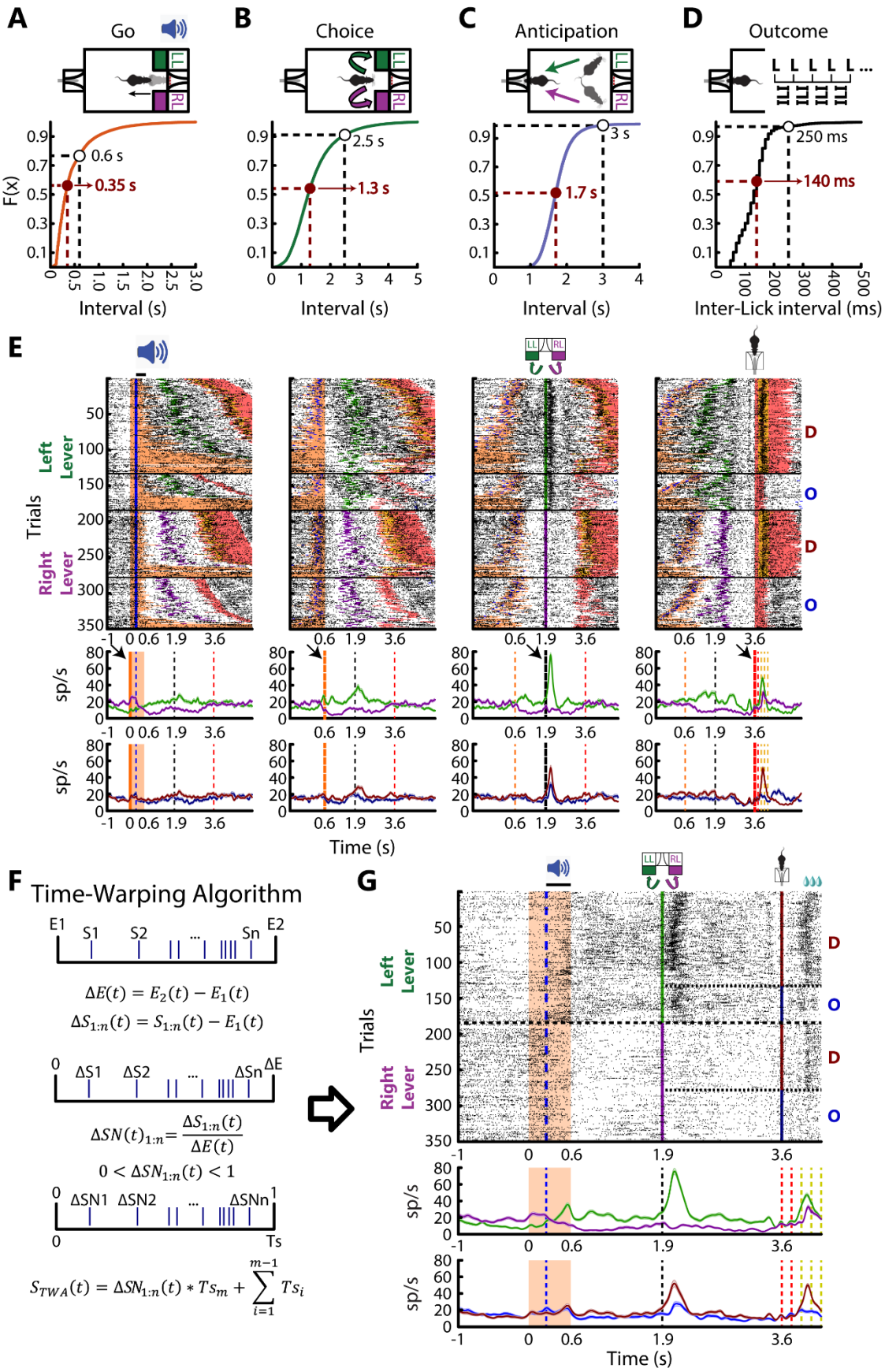

A
Go
LL
RL
F(x)
0.6 s
0.35 s
Interval (s)
B
Choice
2.5 s
1.3 s
Interval (s)
C
Anticipation
3 s
1.7 s
Interval (s)
D
Outcome
250 ms
140 ms
Inter-Lick interval (ms)
E
Left Lever
Right Lever
Trials
D
O
sp/s
Time (s)
F
Time-Warping Algorithm
E1
S1
S2
Sn
E2
$\Delta E(t) = E_2(t) - E_1(t)$
$\Delta S_{1:n}(t) = S_{1:n}(t) - E_1(t)$
0
ΔS1
ΔS2
ΔSn
ΔE
$\Delta SN(t)_{1:n} = \frac{\Delta S_{1:n}(t)}{\Delta E(t)}$
$0 < \Delta SN_{1:n}(t) < 1$
ΔSN1
ΔSN2
ΔSNn
1
Ts
$S_{TWA}(t) = \Delta SN_{1:n}(t) * Ts_m + \sum_{i=1}^{m-1} Ts_i$
G
Left Lever
Right Lever
Trials
D
O
sp/s
Time (s)

**Figure 2. Supplementary 2. Representative raster-plot of a time-warped neuronal response.**

**(A)** Top panel illustrates the Go epoch, the reaction time between the Go signal and exit of the head from the central port. Below, the cumulative distribution function of these time intervals is presented. Dashed red lines show the time constant used for alignment (close to the median of the distribution), with dashed black lines indicating the upper limit (about 1.8 times the median) to prevent an over-shrink effect (see methods).

**(B-C)** Same convention as A, but for Choice and Anticipatory epochs.

**(D)** Depicts the distribution of inter-licks intervals (ILI) observed in the Outcome epoch.

**(E)** A raster plot of an example neuron pre-time-warping. Spiking timestamps (black ticks) of a NAcSh neuron are aligned (time = 0s) to four task epochs: head entry, head out, lever press, and the first empty lick on the outcome spout. The orange-shaded area represents the head-entry time interval in the central port. The vertical blue line indicates the Go signal onset, and the green and purple lines show the left and right lever press times, respectively. The shaded red area denotes the time of licking duration in the reward port. The yellow mark indicates the three drops of sucrose. Below is the PSTH of trials sorted by left/right choices (top PSTHs, green and purple lines, respectively) and reward delivery/omission (bottom PSTHs, red and blue lines, respectively). The orange area depicts the time interval in the central port, the black dashed line indicates the onset of the lever press, the dashed red and yellow lines indicate the times at the first two empty licks and the three licks associated with sucrose were delivered or omitted in the reward port, respectively. The arrows indicate the time for alignment.

**(F)** A graphical depiction of the Time-Warping Algorithm (TWA).

**(G)** A Raster plot of the same neuron is shown in **E** after TWA. This algorithm linearly scales the time intervals between task events on a trial-by-trial basis approximately to the median interval, preserving the number of spikes and overall firing pattern between events.

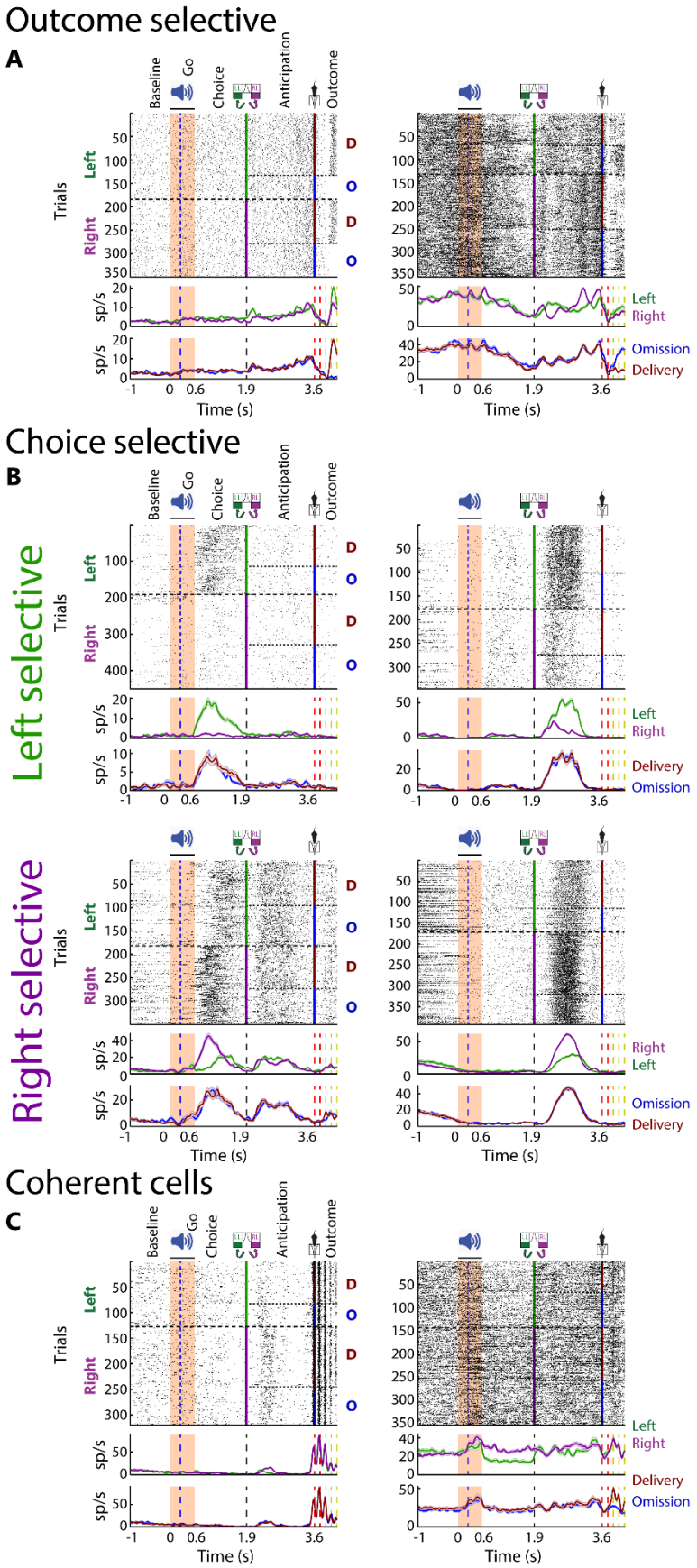

Outcome selective
A
Baseline
Go
Choice
Anticipation
Outcome
Trials
Left
Right
D
O
sp/s
Time (s)
Left
Right
Omission
Delivery
Choice selective
B
Left selective
Right selective
Coherent cells
C

**Figure 2 Supplementary 3. More examples of individual neuronal responses of the NAcSh.**

The neuronal responses are shown with the same convention as in **Figure 2A-C**.

**(A)** NAcSh outcome selective neurons. The left panel depicts a representative neuron that increases its firing rate specifically when a reward (sucrose) is delivered but remains silent when it is absent. In contrast, the right panel shows a neuron that activates only when the reward is omitted.

**(B)** NAcSh choice-selective responses. The upper panels depict two neurons with an increased firing rate selectively during the activation of the left lever. Conversely, the bottom panel presents two neurons whose firing rates become selectively increased in response to the right lever The left panel is a neuron whose firing rates are selective to the respective lever during the Choice epoch, whereas the right panels are selective during the Anticipation epoch.

**(C)** A lick-coherent neuron in the NAcSh. In the left panel, a representative neuron whose firing rate is coherent to licking, while in the right panel, a neuron whose firing rate is coherent only during the sucrose delivery but not omission. Lick-coherent neurons also exhibited a choice-selective response during the Anticipation (left panel) and Choice (right panel) epochs.

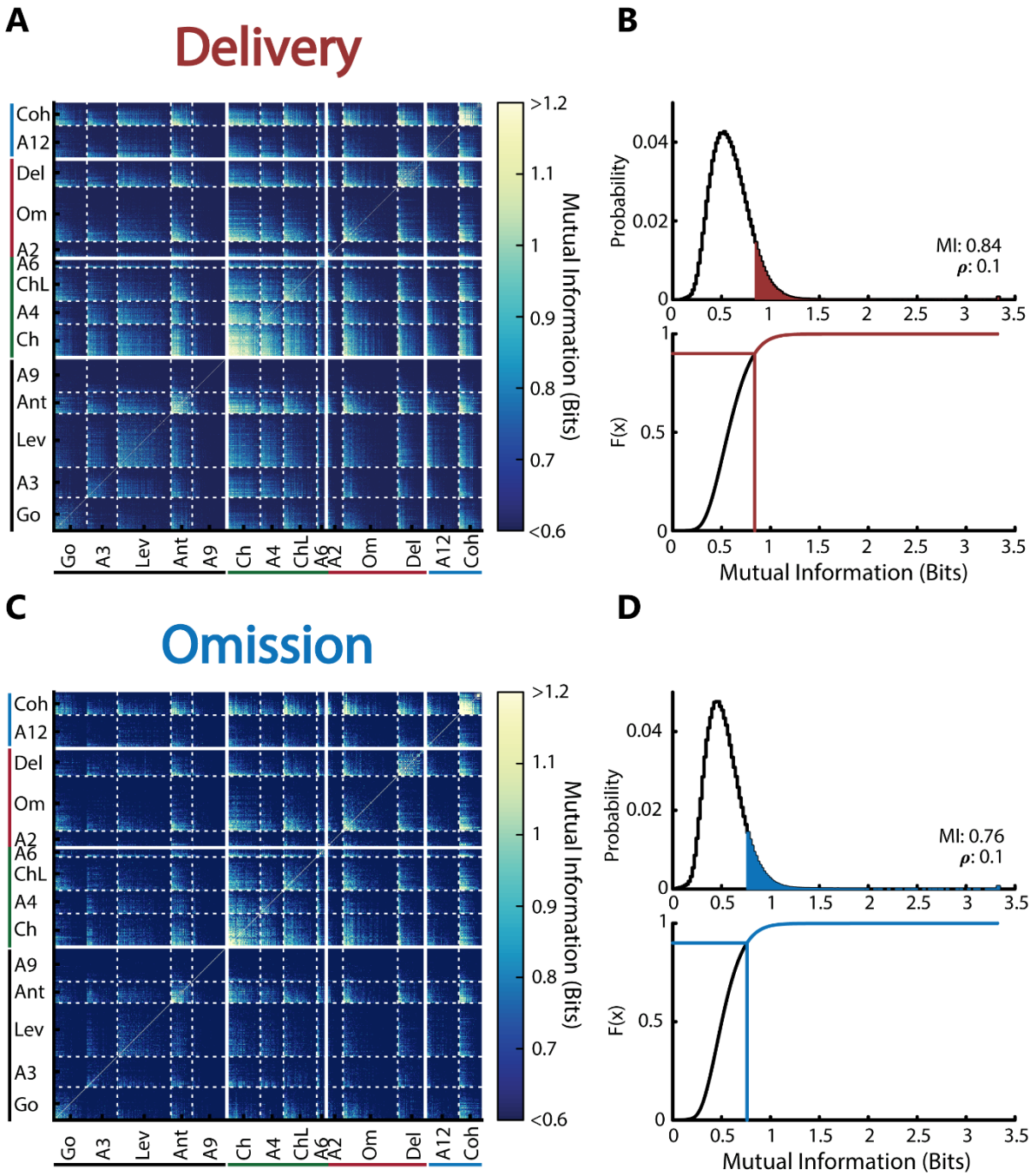



**Figure 5 supplementary 1. Based on the Mutual Information of neuronal activity covariance in the NAcSh, we built the probabilistic network.**

**(A)** The heat map illustrates the mutual information between all neuron connections during reward delivery. The neuronal ensembles sort both axes in the heatmap (**Figure 4**). This metric quantifies the reduction of uncertainty or the degree of statistical dependence that one neuron provides about another and is used to measure the correlation or synchrony

between all pairs of neuronal activity patterns in the NAcSh. The NAcSh is a completely connected network comprising 1507 nodes and 1,134,771 connections. Thus, the "connections" are weighted and undirected.

**(B)** Top panel shows the distribution of the mutual information of all connections, whereas the bottom panel displays the cumulative density function of the same distribution. We consider it "connected" if the mutual information between two neurons exceeds the $90^{th}$ percentile of the mutual information distribution (corresponding to the blue shaded area and the blue line, top and bottom panel, respectively). Otherwise, this was considered as "disconnected."

Panels (**C**) and (**D**) have the same conventions as (**A**) and (**B**) but for reward omission.

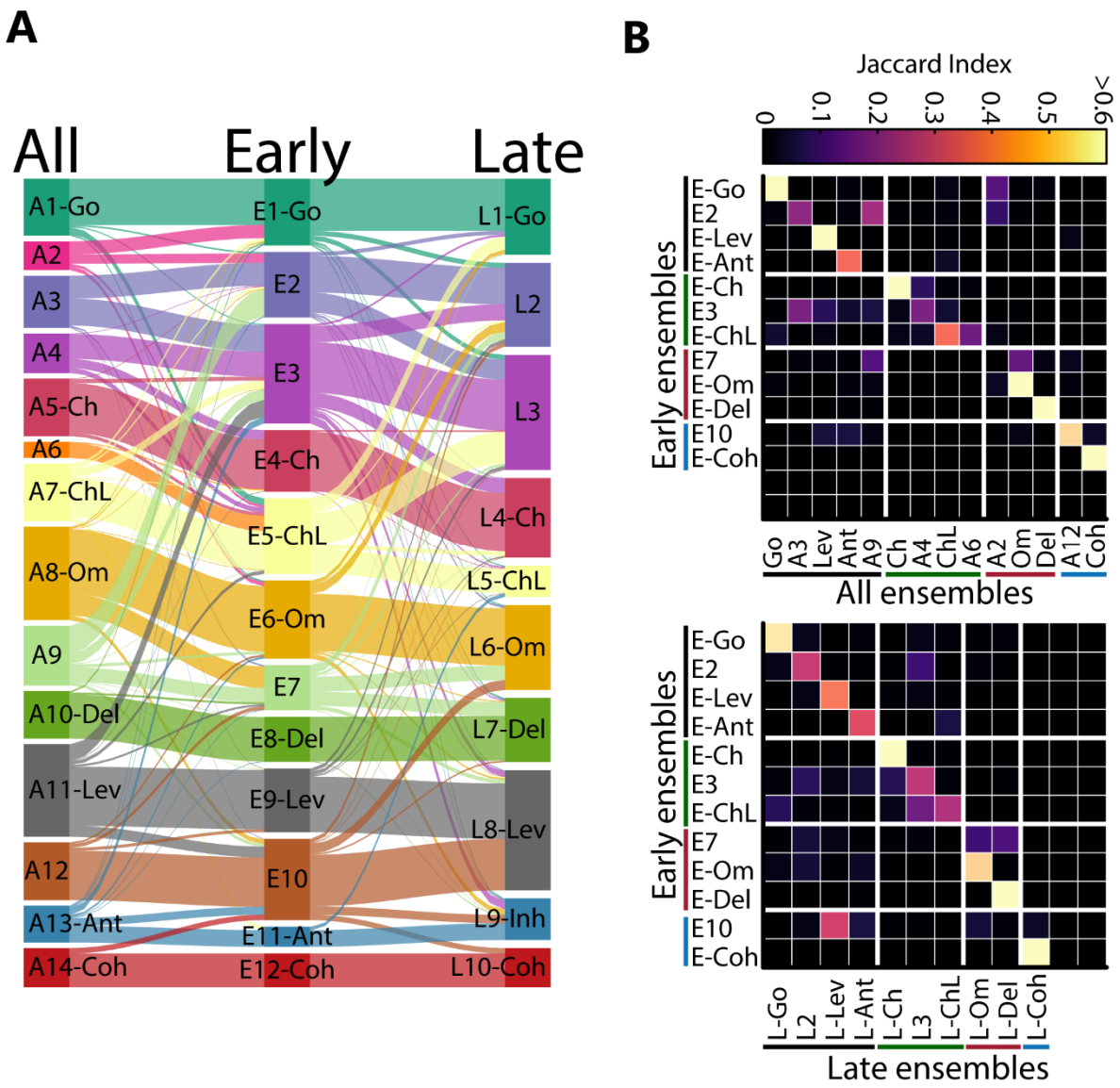



**Figure 7 Supplementary 1. NAcSh neurons drop in and out between neuronal ensembles across All, Early, and Late phases of the block.**

**(A)** Sankey diagram visually represents neuronal ensemble transitions within the block. Note the consistent coloring of ensembles across conditions.

**(B)** Comparison of neuronal ensemble overlapping in the NAcSh ensembles based on t-SNE analysis. The top panel shows the Jaccard index between the neuronal ensembles found in All trials within a block (50 trials) and those found in the Early phase (first 25 trials). The bottom panel displays the overlap between neuronal ensembles in the early and late (last 25 trials) phases. In both comparisons, the overlapping is sorted based on neighbor transitions, highlighting changes in the neuronal ensemble composition across conditions.

Despite a reduction in the number of ensembles identified and by neurons moving between them, the overall transitions can be easily traced using the Jaccard index.

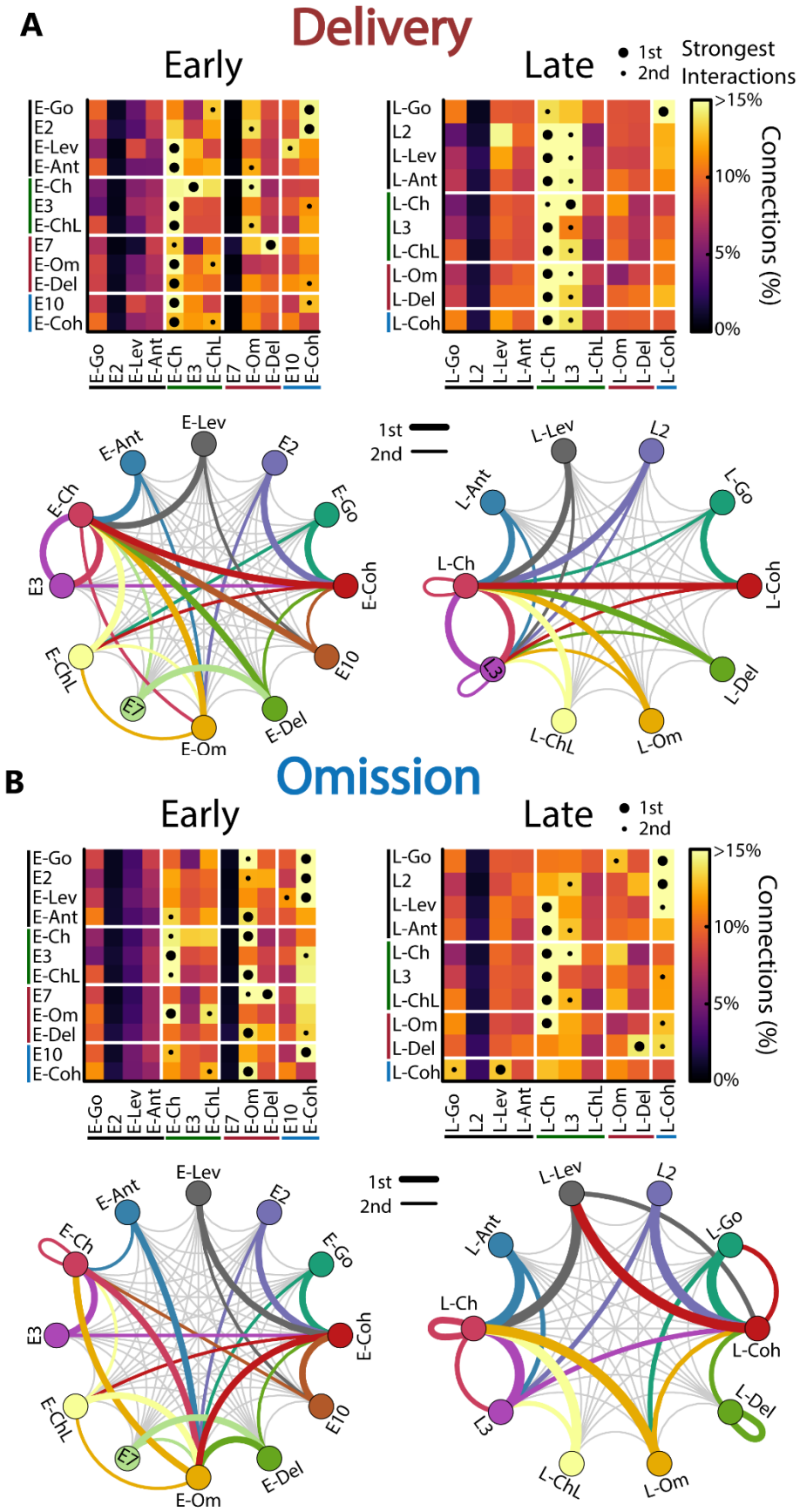

A
Delivery
Early
Late
1st
2nd
Strongest Interactions
>15%
10%
5%
0%
Connections (%)
E-Go
E2
E-Lev
E-Ant
E-Ch
E3
E-ChL
E7
E-Om
E-Del
E10
E-Coh
L-Go
L2
L-Lev
L-Ant
L-Ch
L3
L-ChL
L-Om
L-Del
L-Coh
B
Omission

**Figure 8 Supplementary 1. Delivery and Omission networks during the early and late phases of reinforcement learning**

**(A)** Delivery network reconfiguration during the transition from Early (left) to Late (right) phase. Top panels and heatmaps depict connectivity distribution across ensembles (as in **Figure 5A**) and exhibit the 1st and 2nd strongest interactions. Both axes are sorted based on meta-ensembles Epoch (black line), Choice (green line), Outcome (red line), and Lick Active (blue line). The heatmap is divided as the meta-ensembles by horizontal and vertical solid white lines.

Bottom panels illustrate the Delivery network in a circular layout due to the ensemble's shared connections with all the remaining neuronal ensembles. The 1st and 2nd strongest interactions are indicated by a thick and thin line, respectively. The strongest interactions are colored by the corresponding color assigned to each neuronal ensemble. Gray lines indicate the remaining interactions.

**(B)** Omission network reconfiguration during the transition from Early (left) to Late (right) phase. Same convention as **A**.

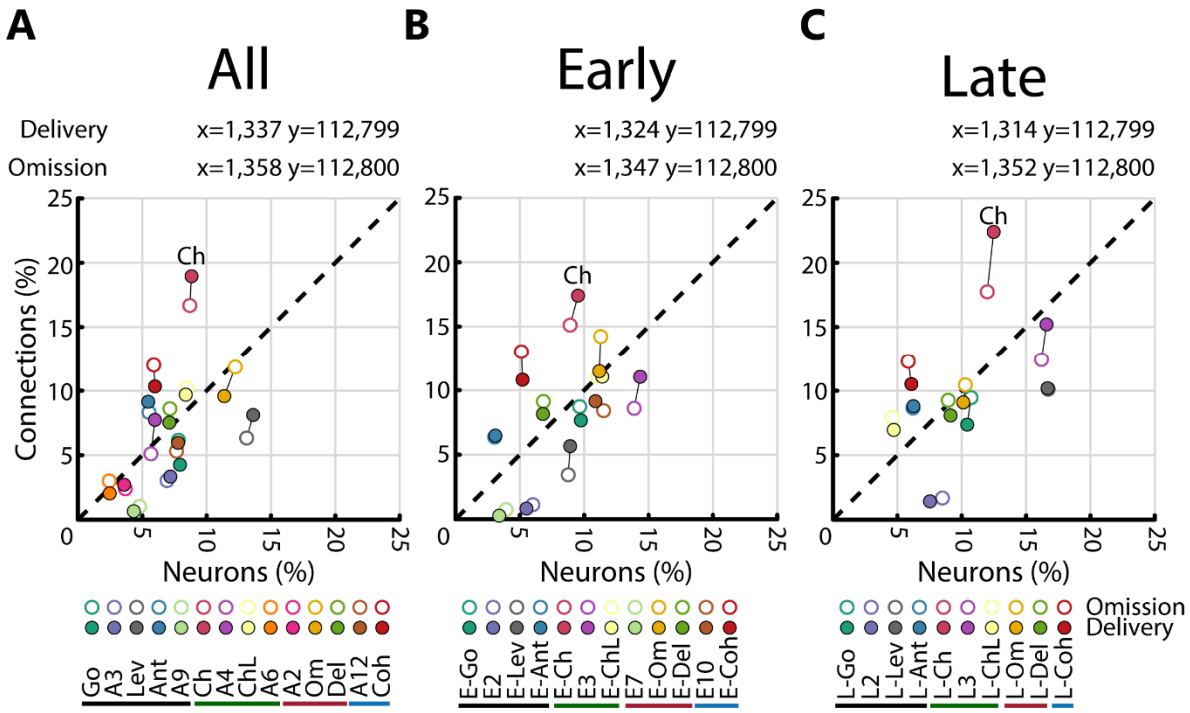



**Figure 8 Supplementary 2. The choice ensemble accounted for the largest number of connections within the NAcSh network.**

**(A-C)** Neuronal Ensembles and Connectivity: These panels explore the relationship between the percentage of neurons and the percentage of connections across neuronal ensembles, both overall (All) and during the transition from Early to Late phases. Each panel plots the number of neurons on the x-axis and connections on the y-axis, with distinct colors representing different ensembles. Filled circles represent reward delivery, while empty circles depict reward omission. The diagonal dashed line indicates equal contribution of neurons and connections.

## Reference


1. Mogenson, G.J., Jones, D.L., and Yim, C.Y. (1980). From motivation to action: Functional interface between the limbic system and the motor system. Progress in Neurobiology *14*, 69–97. 10.1016/0301-0082(80)90018-0.

2. Olds, J. (1956). Pleasure Centers in the Brain. Scientific American. https://www.scientificamerican.com/article/pleasure-centers-in-the-brain/.

3. Floresco, S.B. (2015). The Nucleus Accumbens: An Interface Between Cognition, Emotion, and Action. Annual Review of Psychology *66*, 25–52. 10.1146/annurev-psych-010213-115159.

4. Sugam, J.A., Saddoris, M.P., and Carelli, R.M. (2014). Nucleus Accumbens Neurons Track Behavioral Preferences and Reward Outcomes During Risky Decision Making. Biological Psychiatry *75*, 807–816. 10.1016/j.biopsych.2013.09.010.

5. Kutlu, M.G., Zachry, J.E., Melugin, P.R., Cajigas, S.A., Chevee, M.F., Kelly, S.J., Kutlu, B., Tian, L., Siciliano, C.A., and Calipari, E.S. (2021). Dopamine release in the nucleus accumbens core signals perceived saliency. Curr Biol *31*, 4748-4761.e8. 10.1016/j.cub.2021.08.052.

6. Kelley, A.E., Baldo, B.A., Pratt, W.E., and Will, M.J. (2005). Corticostriatal-hypothalamic circuitry and food motivation: integration of energy, action and reward. Physiol Behav *86*, 773–795. 10.1016/j.physbeh.2005.08.066.

7. Janak, P.H., Chen, M.-T., and Caulder, T. (2004). Dynamics of neural coding in the accumbens during extinction and reinstatement of rewarded behavior. Behavioural Brain Research *154*, 125–135. 10.1016/j.bbr.2004.02.003.

8. Lee, R.-S., Koob, G.F., and Henriksen, S.J. (1998). Electrophysiological responses of nucleus accumbens neurons to novelty stimuli and exploratory behavior in the awake, unrestrained rat. Brain Research *799*, 317–322. 10.1016/S0006-8993(98)00477-6.

9. London, T.D., Licholai, J.A., Szczot, I., Ali, M.A., LeBlanc, K.H., Fobbs, W.C., and Kravitz, A.V. (2018). Coordinated Ramping of Dorsal Striatal Pathways preceding Food Approach and Consumption. J. Neurosci. *38*, 3547–3558. 10.1523/JNEUROSCI.2693-17.2018.

10. Nicola, S.M. (2007). The nucleus accumbens as part of a basal ganglia action selection circuit. Psychopharmacology *191*, 521–550. 10.1007/s00213-006-0510-4.

11. Taha, S.A., and Fields, H.L. (2005). Encoding of Palatability and Appetitive Behaviors by Distinct Neuronal Populations in the Nucleus Accumbens. J. Neurosci. *25*, 1193–1202. 10.1523/JNEUROSCI.3975-04.2005.

12. Tellez, L.A., Perez, I.O., Simon, S.A., and Gutierrez, R. (2012). Transitions between sleep and feeding states in rat ventral striatum neurons. Journal of Neurophysiology *108*, 1739–1751. 10.1152/jn.00394.2012.

13. Wan, X., and Peoples, L.L. (2006). Firing Patterns of Accumbal Neurons During a Pavlovian-Conditioned Approach Task. Journal of Neurophysiology *96*, 652–660. 10.1152/jn.00068.2006.

14. Krause, M., German, P.W., Taha, S.A., and Fields, H.L. (2010). A pause in nucleus accumbens neuron firing is required to initiate and maintain feeding. J Neurosci *30*, 4746–4756. 10.1523/JNEUROSCI.0197-10.2010.

15. O'Connor, E.C., Kremer, Y., Lefort, S., Harada, M., Pascoli, V., Rohner, C., and Lüscher, C. (2015). Accumbal D1R Neurons Projecting to Lateral Hypothalamus Authorize Feeding. Neuron *88*, 553–564. 10.1016/j.neuron.2015.09.038.

16. Prado, L., Luis-Islas, J., Sandoval, O.I., Puron, L., Gil, M.M., Luna, A., Arias-García, M.A., Galarraga, E., Simon, S.A., and Gutierrez, R. (2016). Activation of Glutamatergic Fibers in the Anterior NAc Shell Modulates Reward Activity in the aNAcSh, the Lateral Hypothalamus, and Medial Prefrontal Cortex and Transiently Stops Feeding. J. Neurosci. *36*, 12511–12529. 10.1523/JNEUROSCI.1605-16.2016.

17. Deseyve, C., Domingues, A.V., Carvalho, T.T.A., Armada, G., Correia, R., Vieitas-Gaspar, N., Wezik, M., Pinto, L., Sousa, N., Coimbra, B., et al. (2024). Nucleus accumbens neurons dynamically respond to appetitive and aversive associative learning. Journal of Neurochemistry, jnc.16063. 10.1111/jnc.16063.

18. Pedersen, C.E., Castro, D.C., Gray, M.M., Zhou, Z.C., Piantadosi, S.C., Gowrishankar, R., Kan, S.A., Murphy, P.J., O'Neill, P.R., and Bruchas, M.R. (2022). Medial accumbens shell spiny projection neurons encode relative reward preference. Preprint at bioRxiv, 10.1101/2022.09.18.508426 10.1101/2022.09.18.508426.

19. Reed, S.J., Lafferty, C.K., Mendoza, J.A., Yang, A.K., Davidson, T.J., Grosenick, L., Deisseroth, K., and Britt, J.P. (2018). Coordinated Reductions in Excitatory Input to the Nucleus Accumbens Underlie Food Consumption. Neuron *99*, 1260-1273.e4. 10.1016/j.neuron.2018.07.051.

20. Saulskaya, N.B., and Mikhailova, M.O. (2002). Feeding-induced decrease in extracellular glutamate level in the rat nucleus accumbens: dependence on glutamate uptake. Neuroscience *112*, 791–801. 10.1016/S0306-4522(02)00126-4.

21. Britt, J.P., Benaliouad, F., McDevitt, R.A., Stuber, G.D., Wise, R.A., and Bonci, A. (2012). Synaptic and behavioral profile of multiple glutamatergic inputs to the nucleus accumbens. Neuron *76*, 790–803. 10.1016/j.neuron.2012.09.040.

22. Coss, A., Suaste, E., and Gutierrez, R. (2022). Lateral NAc Shell D1 and D2 Neuronal Ensembles Concurrently Predict Licking Behavior and Categorize Sucrose Concentrations in a Context-dependent Manner. Neuroscience *493*, 81–98. 10.1016/j.neuroscience.2022.04.022.

23. Gutierrez, R., Simon, S.A., and Nicolelis, M.A.L. (2010). Licking-Induced Synchrony in the Taste–Reward Circuit Improves Cue Discrimination during Learning. J. Neurosci. *30*, 287–303. 10.1523/JNEUROSCI.0855-09.2010.

24. Villavicencio, M., Moreno, M.G., Simon, S.A., and Gutierrez, R. (2018). Encoding of Sucrose's Palatability in the Nucleus Accumbens Shell and Its Modulation by Exteroceptive Auditory Cues. Frontiers in Neuroscience *12*.

25. Carrillo-Reid, L., and Yuste, R. (2020). What Is a Neuronal Ensemble? In Oxford Research Encyclopedia of Neuroscience 10.1093/acrefore/9780190264086.013.298.

26. Lara-González, E., Padilla-Orozco, M., Fuentes-Serrano, A., Bargas, J., and Duhne, M. (2022). Translational neuronal ensembles: Neuronal microcircuits in psychology, physiology, pharmacology and pathology. Frontiers in Systems Neuroscience *16*.

27. Yuste, R., Cossart, R., and Yaksi, E. (2024). Neuronal ensembles: Building blocks of neural circuits. Neuron. 10.1016/j.neuron.2023.12.008.

28. Hamid, A.A., Pettibone, J.R., Mabrouk, O.S., Hetrick, V.L., Schmidt, R., Vander Weele, C.M., Kennedy, R.T., Aragona, B.J., and Berke, J.D. (2016). Mesolimbic dopamine signals the value of work. Nat Neurosci *19*, 117–126. 10.1038/nn.4173.

29. Faust, T.W., Mohebi, A., and Berke, J.D. (2023). Reward expectation selectively boosts the firing of accumbens D1+ neurons during motivated approach. Preprint at bioRxiv, 10.1101/2023.09.02.556060 10.1101/2023.09.02.556060.

30. Swanson, K., Averbeck, B.B., and Laubach, M. (2022). Noradrenergic regulation of two-armed bandit performance. Behavioral Neuroscience *136*, 84–99. 10.1037/bne0000495.

31. Stalnaker, T.A., Berg, B., Aujla, N., and Schoenbaum, G. (2016). Cholinergic Interneurons Use Orbitofrontal Input to Track Beliefs about Current State. J. Neurosci. *36*, 6242–6257. 10.1523/JNEUROSCI.0157-16.2016.

32. Chantranupong, L., Beron, C.C., Zimmer, J.A., Wen, M.J., Wang, W., and Sabatini, B.L. (2023). Dopamine and glutamate regulate striatal acetylcholine in decision-making. Nature *621*, 577–585. 10.1038/s41586-023-06492-9.

33. Williams, A.H., Poole, B., Maheswaranathan, N., Dhawale, A.K., Fisher, T., Wilson, C.D., Brann, D.H., Trautmann, E.M., Ryu, S., Shusterman, R., et al. (2020). Discovering Precise Temporal Patterns in Large-Scale Neural Recordings through Robust and Interpretable Time Warping. Neuron *105*, 246-259.e8. 10.1016/j.neuron.2019.10.020.

34. Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., and Movshon, J.A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis Neurosci *13*, 87–100. 10.1017/s095252380000715x.

35. Chung, S., and Abbott, L.F. (2021). Neural population geometry: An approach for understanding biological and artificial neural networks. Current Opinion in Neurobiology *70*, 137–144. 10.1016/j.conb.2021.10.010.

36. Shinn, M. (2023). Phantom oscillations in principal component analysis. Proceedings of the National Academy of Sciences *120*, e2311420120. 10.1073/pnas.2311420120.

37. Laboratory, I.B., Benson, B., Benson, J., Birman, D., Bonacchi, N., Carandini, M., Catarino, J.A., Chapuis, G.A., Churchland, A.K., Dan, Y., et al. (2023). A Brain-Wide Map of Neural Activity during Complex Behaviour. Preprint at bioRxiv, 10.1101/2023.07.04.547681 10.1101/2023.07.04.547681.

38. Watts, D.J., and Strogatz, S.H. (1998). Collective dynamics of “small-world” networks. Nature *393*, 440–442. 10.1038/30918.

39. Barabasi, A.L., and Albert, R. (1999). Emergence of scaling in random networks. Science *286*, 509–512. 10.1126/science.286.5439.509.

40. Ito, M., and Doya, K. (2015). Distinct Neural Representation in the Dorsolateral, Dorsomedial, and Ventral Parts of the Striatum during Fixed- and Free-Choice Tasks. Journal of Neuroscience *35*, 3499–3514. 10.1523/JNEUROSCI.1962-14.2015.

41. Sutton, R.S., and Barto, A.G. (2018). Reinforcement Learning: An Introduction (A Bradford Book).

42. Beroun, A., Bijoch, Ł., Pawlowska, M., Wiśniewska, J., Legutko, D., Szachowicz, U., and Kaczmarek, L. (2022). Whole-brain tracking of addictive and natural rewards processing. 10.21203/rs.3.rs-2185874/v1.

43. Münte, T., Heldmann, M., Hinrichs, H., Marco-Pallares, J., Krämer, U., Sturm, V., and Heinze, H.-J. (2008). Nucleus accumbens is involved in human action monitoring: evidence from invasive electrophysiological recordings. Frontiers in Human Neuroscience *2*.

44. O’Donnell, P. (1999). Ensemble coding in the nucleus accumbens. Psychobiology *27*, 187–197. 10.3758/BF03332113.

45. Roitman, M.F., Wheeler, R.A., and Carelli, R.M. (2005). Nucleus Accumbens Neurons Are Innately Tuned for Rewarding and Aversive Taste Stimuli, Encode Their Predictors, and Are Linked to Motor Output. Neuron *45*, 587–597. 10.1016/j.neuron.2004.12.055.

46. Volman, S.F., Lammel, S., Margolis, E.B., Kim, Y., Richard, J.M., Roitman, M.F., and Lobo, M.K. (2013). New insights into the specificity and plasticity of reward and aversion encoding in the mesolimbic system. J Neurosci *33*, 17569–17576. 10.1523/JNEUROSCI.3250-13.2013.

47. Chen, S., Liu, Y., Wang, Z.A., Colonell, J., Liu, L.D., Hou, H., Tien, N.-W., Wang, T., Harris, T., Druckmann, S., et al. (2024). Brain-wide neural activity underlying memory-guided movement. Cell *187*, 676-691.e16. 10.1016/j.cell.2023.12.035.

48. Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., and Churchland, A.K. (2019). Single-trial neural dynamics are dominated by richly varied movements. Nat Neurosci *22*, 1677–1686. 10.1038/s41593-019-0502-4.

49. Steinmetz, N.A., Zatka-Haas, P., Carandini, M., and Harris, K.D. (2019). Distributed coding of choice, action and engagement across the mouse brain. Nature *576*, 266–273. 10.1038/s41586-019-1787-x.

50. Zhou, S., Masmanidis, S.C., and Buonomano, D.V. (2020). Neural Sequences as an Optimal Dynamical Regime for the Readout of Time. Neuron *108*, 651-658.e5. 10.1016/j.neuron.2020.08.020.

51. Fonseca, E., de Lafuente, V., Simon, S.A., and Gutierrez, R. (2018). Sucrose intensity coding and decision-making in rat gustatory cortices. eLife *7*, e41152. 10.7554/eLife.41152.

52. Gutierrez, R., and Simon, S.A. (2021). Physiology of Taste Processing in the Tongue, Gut, and Brain. Compr Physiol *11*, 2489–2523. 10.1002/cphy.c210002.

53. Pennartz, C.M., Groenewegen, H.J., and Lopes da Silva, F.H. (1994). The nucleus accumbens as a complex of functionally distinct neuronal ensembles: an integration of behavioural, electrophysiological and anatomical data. Prog Neurobiol *42*, 719–761. 10.1016/0301-0082(94)90025-6.

54. Bernklau, T.W., Righetti, B., Mehrke, L.S., and Jacob, S.N. (2024). Striatal dopamine signals reflect perceived cue-action-outcome associations in mice. Nat Neurosci. 10.1038/s41593-023-01567-2.

55. Lynn, C.W., Holmes, C.M., and Palmer, S.E. (2024). Heavy-tailed neuronal connectivity arises from Hebbian self-organization. Nat. Phys., 1–8. 10.1038/s41567-023-02332-9.

56. Bargas, J., and Pérez-Ortega, J. (2017). Microcircuits neuroscience to understand pathophysiology. Oncotarget *8*, 7226–7227. 10.18632/oncotarget.14792.

57. Bouaichi, C.G., and Vincis, R. (2020). Cortical processing of chemosensory and hedonic features of taste in active licking mice. Journal of Neurophysiology *123*, 1995–2009. 10.1152/jn.00069.2020.

58. Tomé, D.F., Zhang, Y., Aida, T., Mosto, O., Lu, Y., Chen, M., Sadeh, S., Roy, D.S., and Clopath, C. (2024). Dynamic and selective engrams emerge with memory consolidation. Nat Neurosci, 1–12. 10.1038/s41593-023-01551-w.

59. Yao, Z., van Velthoven, C.T.J., Kunst, M., Zhang, M., McMillen, D., Lee, C., Jung, W., Goldy, J., Abdelhak, A., Aitken, M., et al. (2023). A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. Nature *624*, 317–332. 10.1038/s41586-023-06812-z.

60. Kim, D., Jeong, H., Lee, J., Ghim, J.-W., Her, E.S., Lee, S.-H., and Jung, M.W. (2016). Distinct Roles of Parvalbumin- and Somatostatin-Expressing Interneurons in Working Memory. Neuron *92*, 902–915. 10.1016/j.neuron.2016.09.023.

61. Guenthner, C.J., Miyamichi, K., Yang, H.H., Heller, H.C., and Luo, L. (2013). Permanent genetic access to transiently active neurons via TRAP: targeted recombination in active populations. Neuron *78*, 773–784. 10.1016/j.neuron.2013.03.025.

62. Lee, D., Hyun, J.H., Jung, K., Hannan, P., and Kwon, H.-B. (2017). A calcium- and light-gated switch to induce gene expression in activated neurons. Nat Biotechnol *35*, 858–863. 10.1038/nbt.3902.

63. Stackmann, M., Yelhekar, T.D., Meng, M., Sun, X., Nthumba, J., Bulthuis, N.E., Vaughan, N., Zhu, E., Lin, Y., and Denny, C.A. (2024). A multiple Arc (mArc) tagging system to uncover the organizational principles of multiple memories. Preprint at bioRxiv, 10.1101/2024.02.01.578410 10.1101/2024.02.01.578410.

64. Carrillo-Reid, L., Han, S., Yang, W., Akrouh, A., and Yuste, R. (2019). Controlling Visually Guided Behavior by Holographic Recalling of Cortical Ensembles. Cell *178*, 447-457.e5. 10.1016/j.cell.2019.05.045.

65. Josselyn, S.A., Köhler, S., and Frankland, P.W. (2015). Finding the engram. Nat Rev Neurosci *16*, 521–534. 10.1038/nrn4000.

66. Carrillo-Reid, L., Yang, W., Bando, Y., Peterka, D.S., and Yuste, R. (2016). Imprinting and recalling cortical ensembles. Science *353*, 691–694. 10.1126/science.aaf7560.

67. Pérez-Ortega, J., Duhne, M., Lara-González, E., Plata, V., Gasca, D., Galarraga, E., Hernández-Cruz, A., and Bargas, J. (2016). Pathophysiological signatures of functional connectomics in parkinsonian and dyskinetic striatal microcircuits. Neurobiology of Disease *91*, 347–361. 10.1016/j.nbd.2016.02.023.

68. Chen, R., Blosser, T.R., Djekidel, M.N., Hao, J., Bhattacherjee, A., Chen, W., Tuesta, L.M., Zhuang, X., and Zhang, Y. (2021). Decoding molecular and cellular heterogeneity of mouse nucleus accumbens. Nat Neurosci *24*, 1757–1771. 10.1038/s41593-021-00938-x.

69. McInnes, L., Healy, J., and Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at arXiv, 10.48550/arXiv.1802.03426 10.48550/arXiv.1802.03426.

70. Schneider, S., Lee, J.H., and Mathis, M.W. (2023). Learnable latent embeddings for joint behavioural and neural analysis. Nature *617*, 360–368. 10.1038/s41586-023-06031-6.

71. Song, Z., Alpers, A., Warner, K., Iacobucci, F., Hoskins, E., Disterhoft, J.F., Voss, J.L., and Widge, A.S. (2024). Chronic, Reusable, Multiday Neuropixels Recordings during Free-Moving Operant Behavior. eNeuro *11*. 10.1523/ENEURO.0245-23.2023.

72. Gutierrez, R. (2023). Studying behavior under constrained movement. eLife *12*, e91145. 10.7554/eLife.91145.

73. Elber-Dorozko, L., and Loewenstein, Y. (2018). Striatal action-value neurons reconsidered. eLife *7*, e34248. 10.7554/eLife.34248.

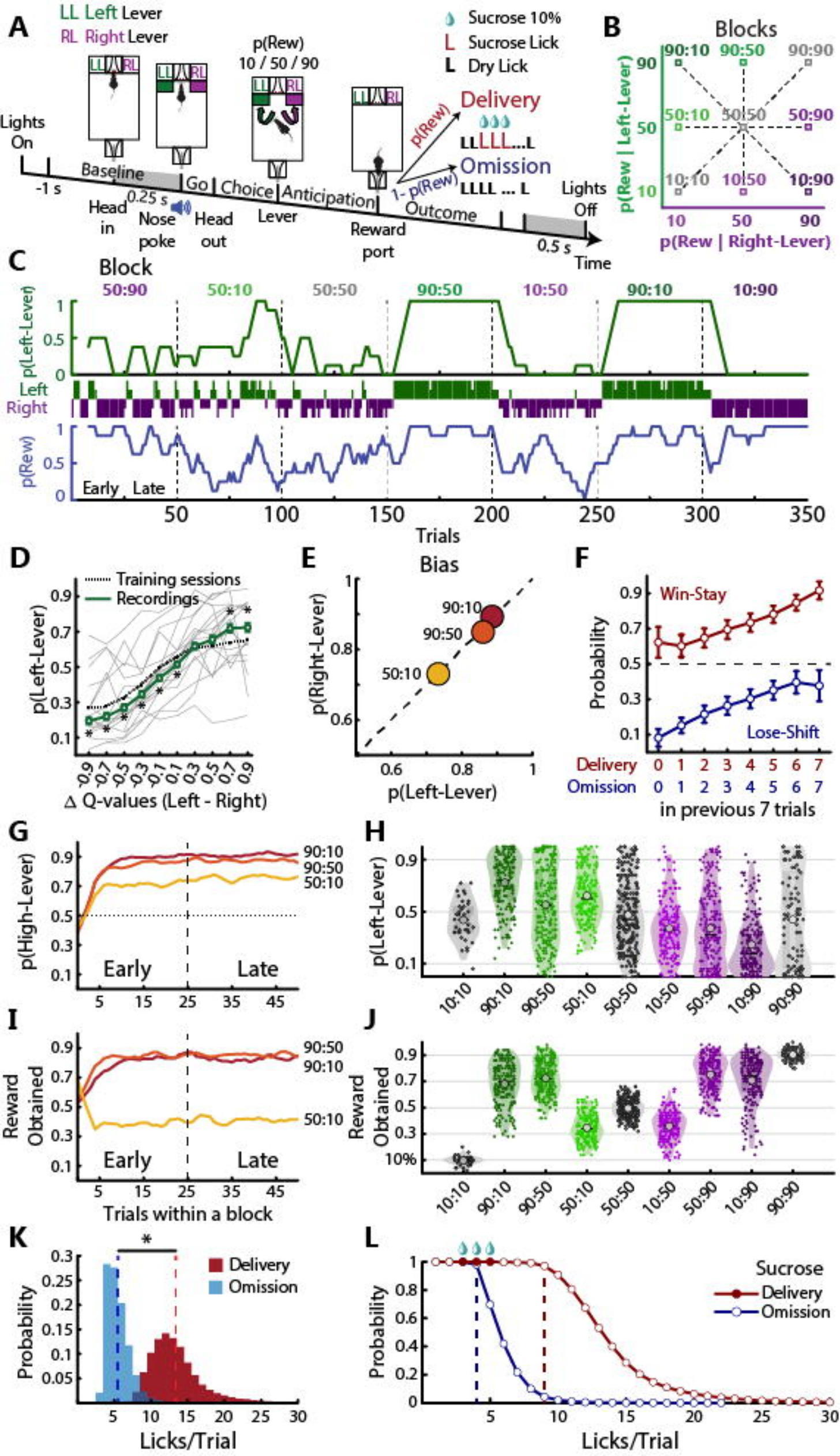
74. Jarvis, M.R., and Mitra, P.P. (2001). Sampling properties of the spectrum and coherency of sequences of action potentials. Neural Comput *13*, 717–749. 10.1162/089976601300014312.

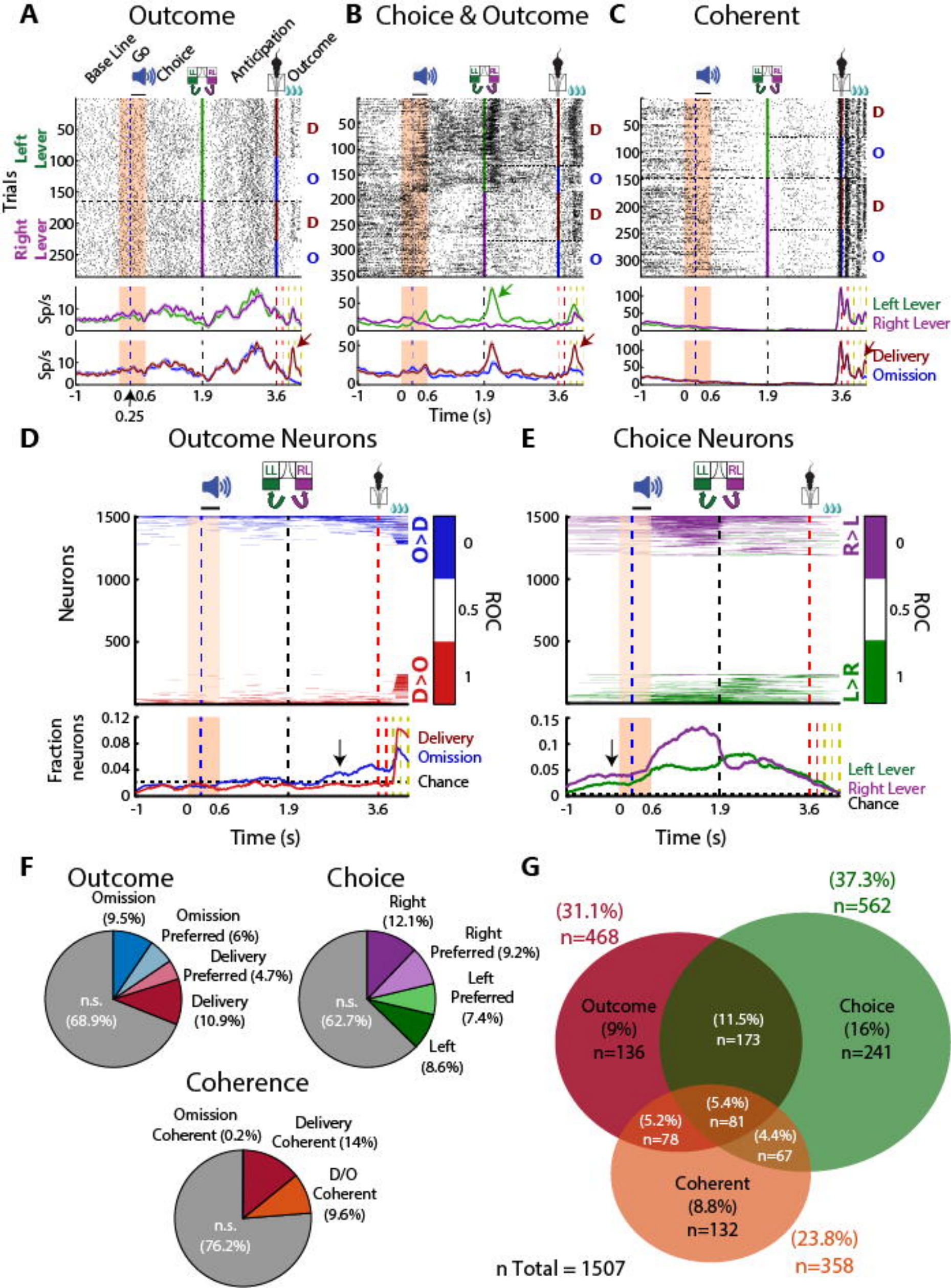
75. Maaten, L. van der, and Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research *9*, 2579–2605.

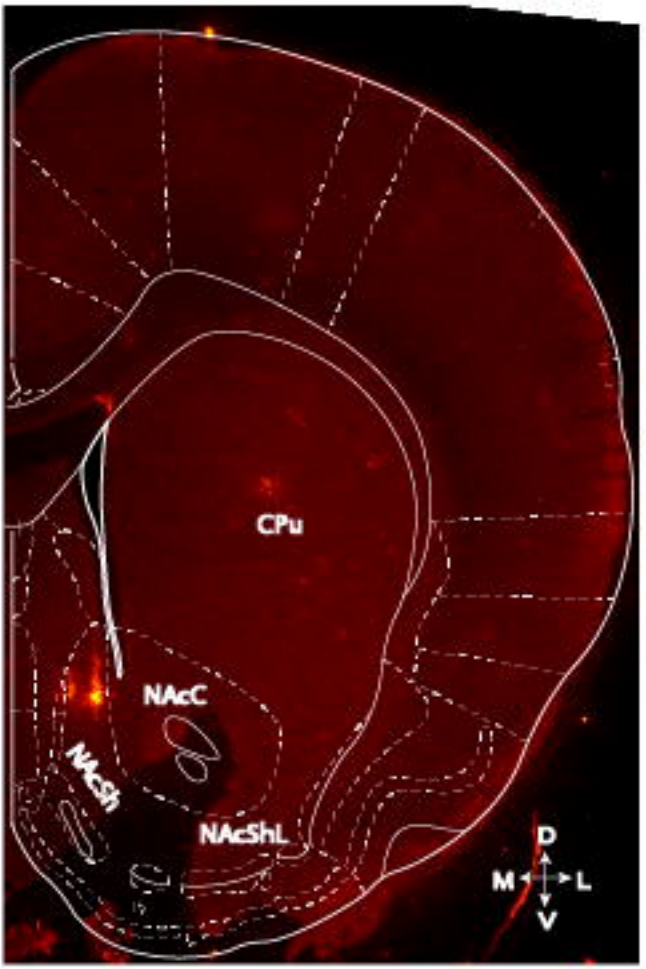
76. Bellot, P., and Meyer, P.E. (2017). Efficient Combination of Pairwise Feature Networks. In Neural Connectomics Challenge The Springer Series on Challenges in Machine Learning., D.

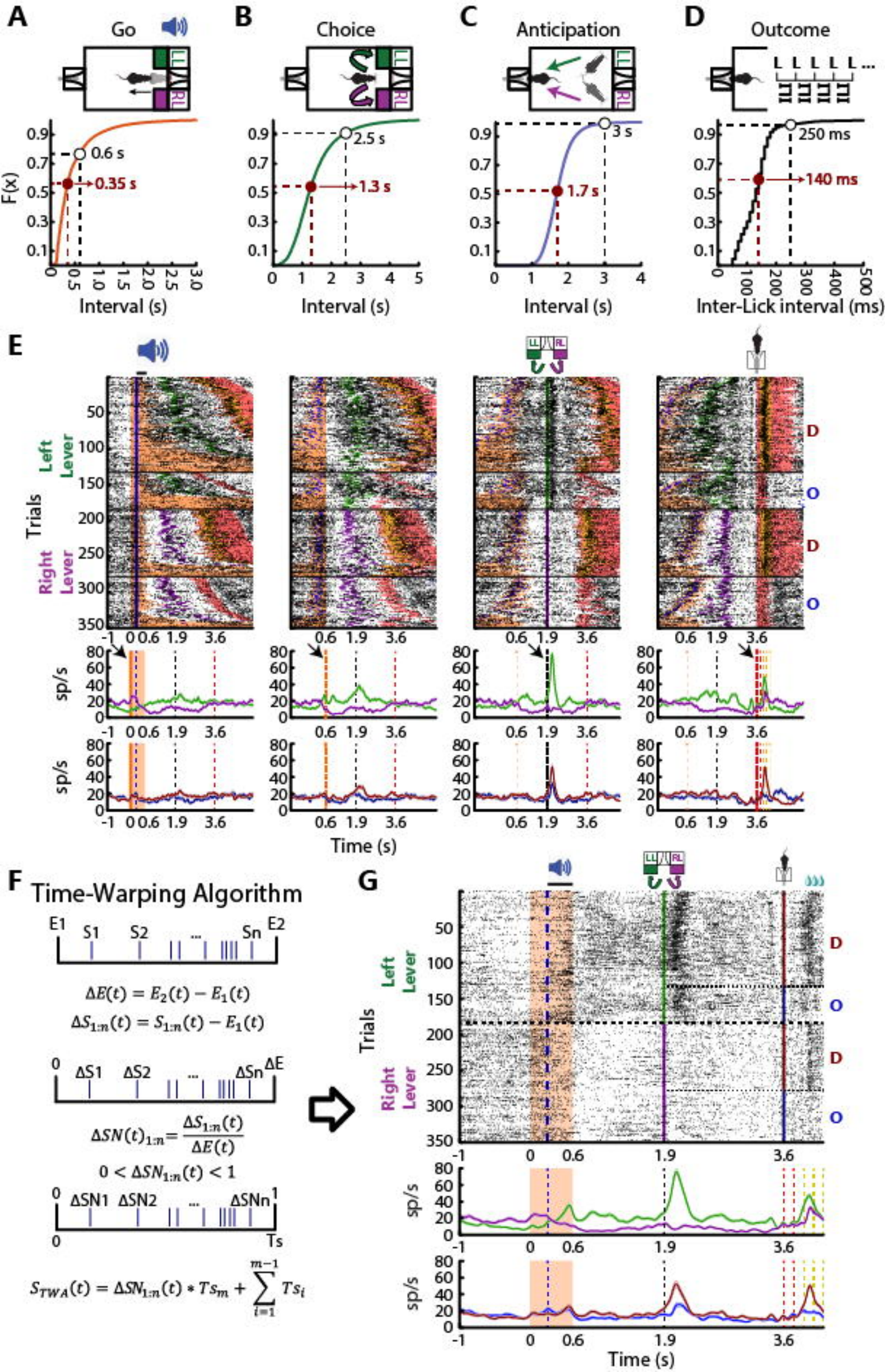
Battaglia, I. Guyon, V. Lemaire, J. Orlandi, B. Ray, and J. Soriano, eds. (Springer International Publishing), pp. 85–93. 10.1007/978-3-319-53070-3_7.

77. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498–2504. 10.1101/gr.1239303.


A
LL Left Lever
RL Right Lever
p(Rew)
10 / 50 / 90
Sucrose 10%
L Sucrose Lick
L Dry Lick
Delivery
LLLLL...L
Omission
LLLL ... L
p(Rew)
1- p(Rew)
Lights On
-1 s
Baseline
Head in
0.25 s
Nose poke
Head out
Go
Choice
Lever
Anticipation
Reward port
Outcome
0.5 s
Lights Off
Time
B
Blocks
90:10
90:50
90:90
50:10
50:50
50:90
10:10
10:50
10:90
p(Rew | Left-Lever)
p(Rew | Right-Lever)
C
Block
50:90
50:10
50:50
90:50
10:50
90:10
10:90
p(Left-Lever)
Left
Right
p(Rew)
Early
Late
Trials
D
Training sessions
Recordings
p(Left-Lever)
Δ Q-values (Left - Right)
E
Bias
90:10
90:50
50:10
p(Right-Lever)
p(Left-Lever)
F
Win-Stay
Lose-Shift
Probability
Delivery 0 1 2 3 4 5 6 7
Omission 0 1 2 3 4 5 6 7
in previous 7 trials
G
p(High-Lever)
90:10
90:50
50:10
Early
Late
H
p(Left-Lever)
10:10
90:10
90:50
50:10
50:50
10:50
50:90
10:90
90:90
I
Reward Obtained
90:50
90:10
50:10
Early
Late
Trials within a block
J
Reward Obtained
10%
K
Delivery
Omission
Probability
Licks/Trial
L
Sucrose
Delivery
Omission
Probability
Licks/Trial

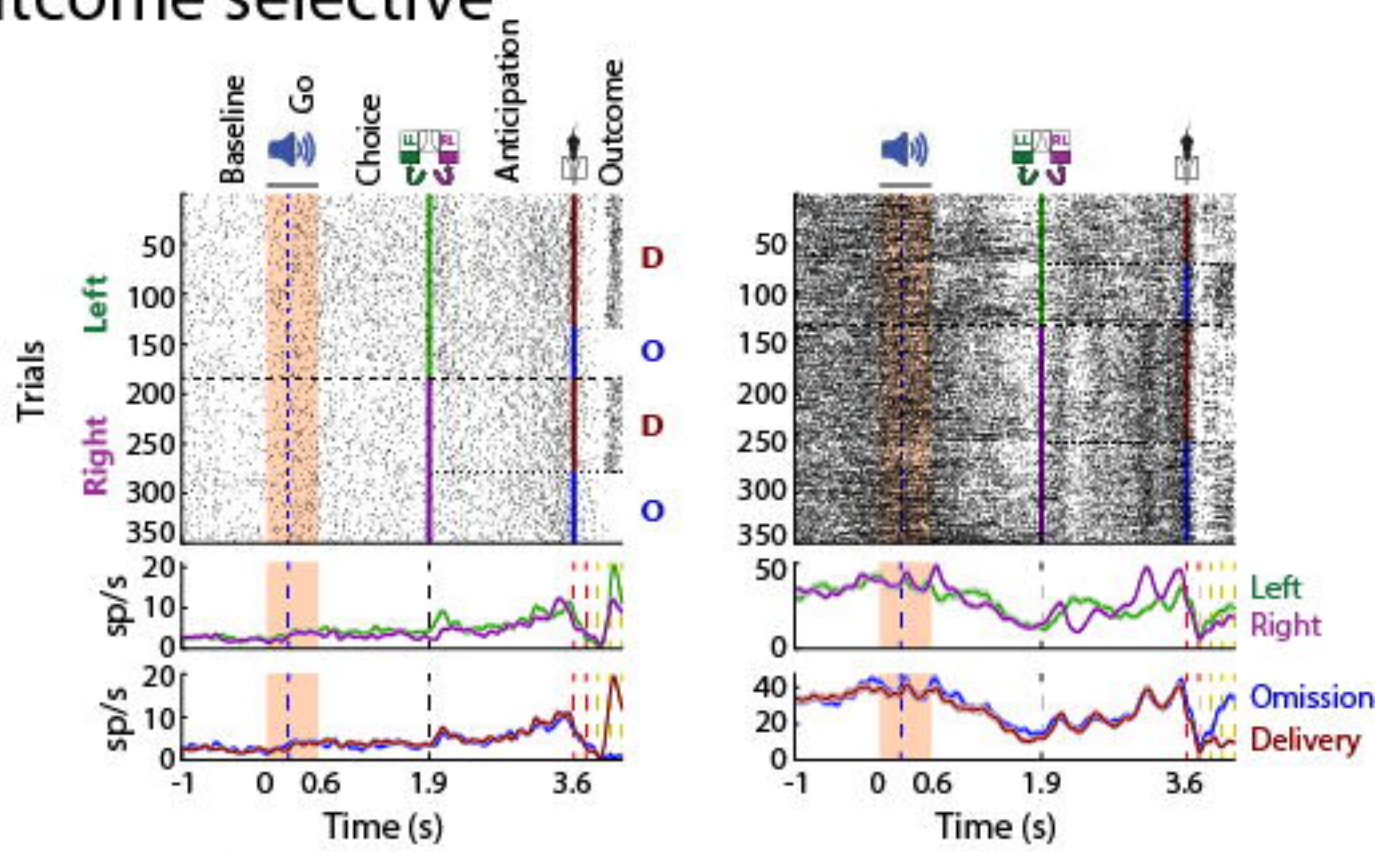

A
Outcome
B
Choice & Outcome
C
Coherent
Base Line
Go
Choice
Anticipation
Outcome
Trials
Left Lever
Right Lever
D
O
Sp/s
Left Lever
Right Lever
Delivery
Omission
Time (s)
D
Outcome Neurons
E
Choice Neurons
Neurons
O>D
D>O
R>L
L>R
ROC
Fraction neurons
Delivery
Omission
Chance
Left Lever
Right Lever
Chance
F
Outcome
Omission (9.5%)
Omission Preferred (6%)
Delivery Preferred (4.7%)
Delivery (10.9%)
n.s. (68.9%)
Choice
Right (12.1%)
Right Preferred (9.2%)
Left Preferred (7.4%)
Left (8.6%)
n.s. (62.7%)
Coherence
Omission Coherent (0.2%)
Delivery Coherent (14%)
D/O Coherent (9.6%)
n.s. (76.2%)
G
(31.1%) n=468
(37.3%) n=562
Outcome (9%) n=136
(11.5%) n=173
Choice (16%) n=241
(5.4%) n=81
(5.2%) n=78
(4.4%) n=67
Coherent (8.8%) n=132
(23.8%) n=358
n Total = 1507


CPu
NAcC
NAcSh
NAcShL
D
M
L
V


A
Go
LL
RL
F(x)
0.6 s
0.35 s
Interval (s)
B
Choice
2.5 s
1.3 s
Interval (s)
C
Anticipation
3 s
1.7 s
Interval (s)
D
Outcome
L L L L L ...
250 ms
140 ms
Inter-Lick interval (ms)
E
Trials
Left Lever
Right Lever
D
O
sp/s
Time (s)
F
Time-Warping Algorithm
E1
S1
S2
...
Sn
E2
$\Delta E(t) = E_2(t) - E_1(t)$
$\Delta S_{1:n}(t) = S_{1:n}(t) - E_1(t)$
0
ΔS1
ΔS2
ΔSn
ΔE
$\Delta SN(t)_{1:n} = \frac{\Delta S_{1:n}(t)}{\Delta E(t)}$
$0 < \Delta SN_{1:n}(t) < 1$
ΔSN1
ΔSN2
ΔSNn
1
Ts
$S_{TWA}(t) = \Delta SN_{1:n}(t) * Ts_m + \sum_{i=1}^{m-1} Ts_i$
G
Trials
Left Lever
Right Lever
D
O
sp/s
Time (s)

# Outcome selective

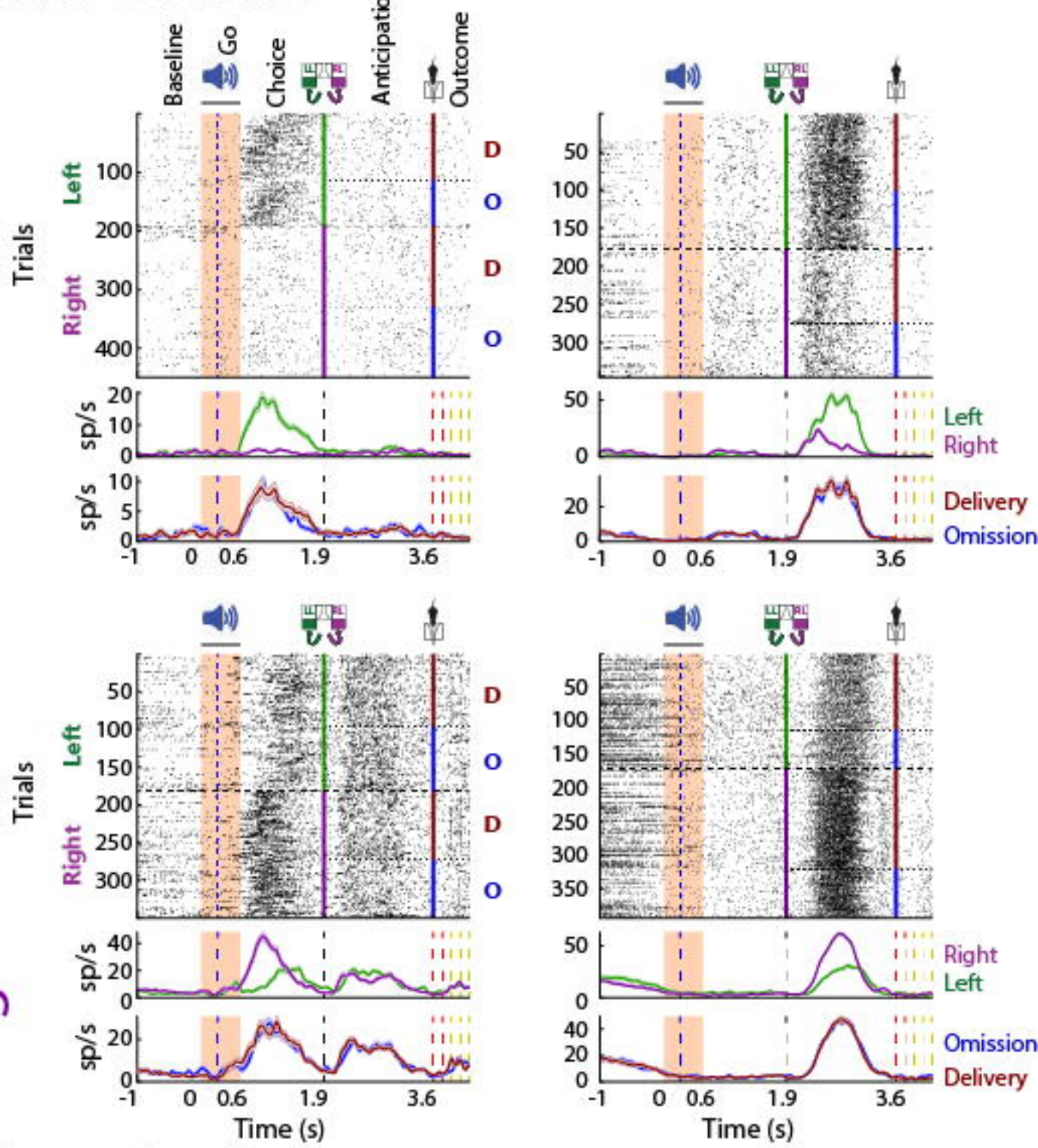
A




# Choice selective

B

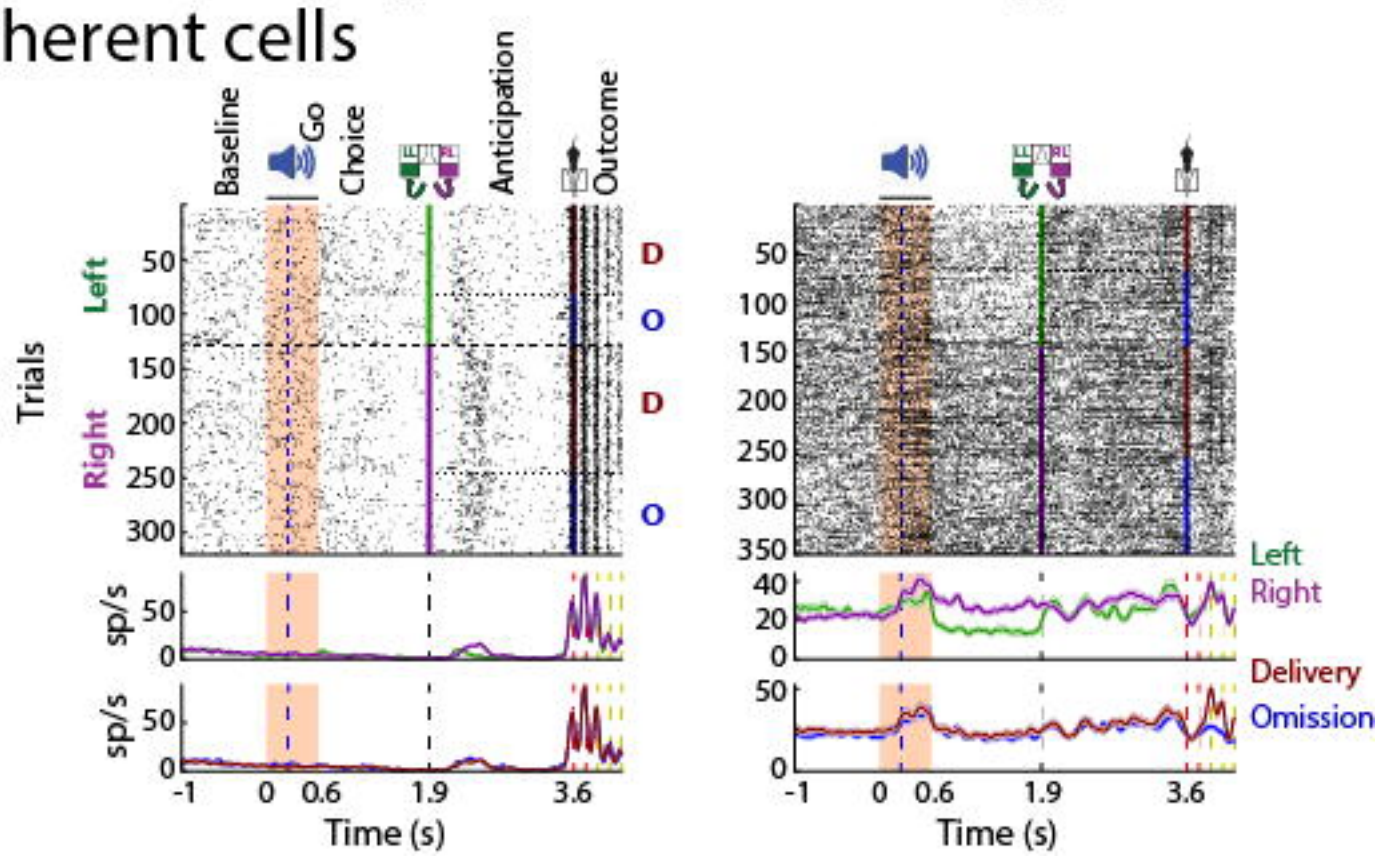
Left selective

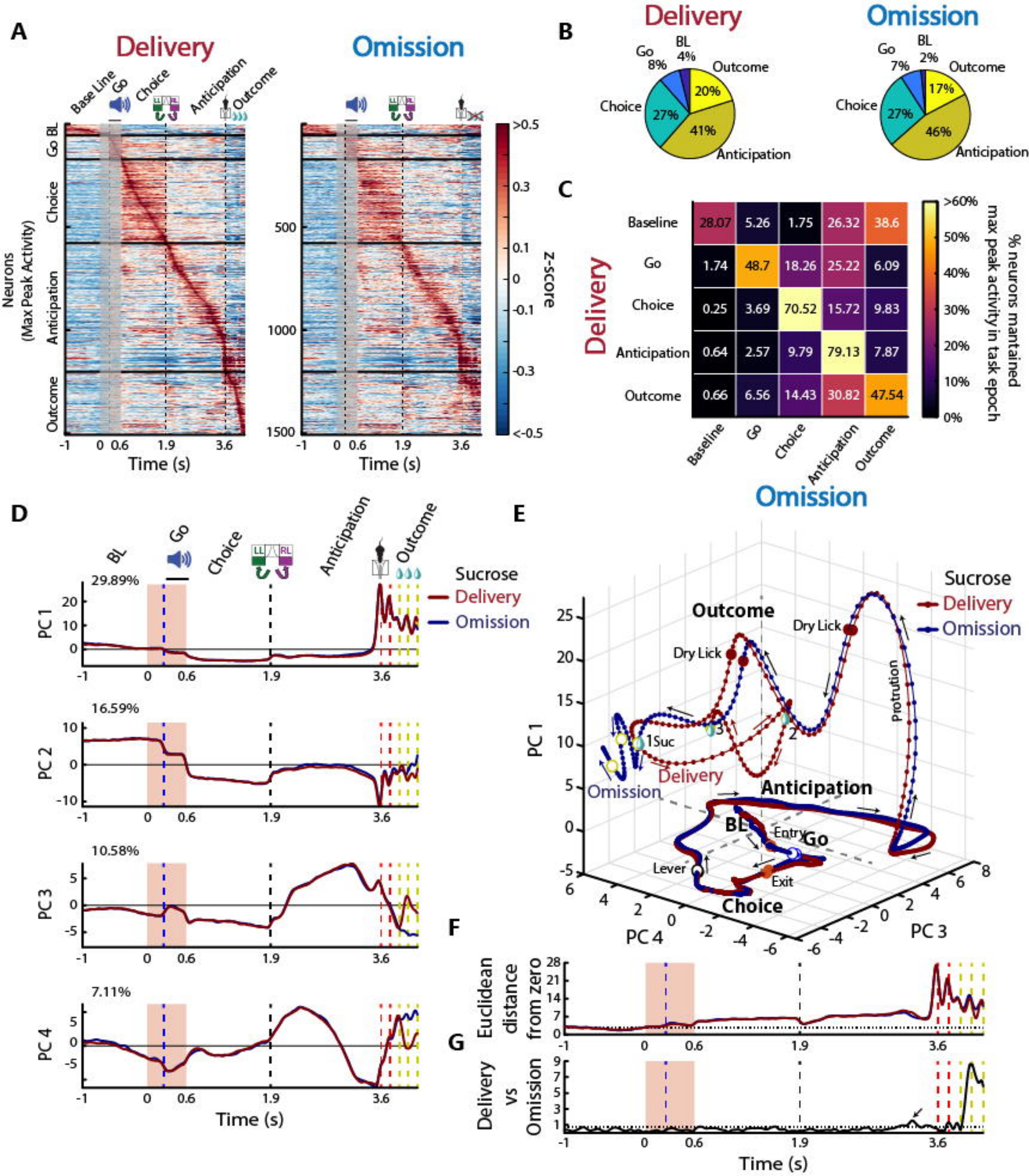


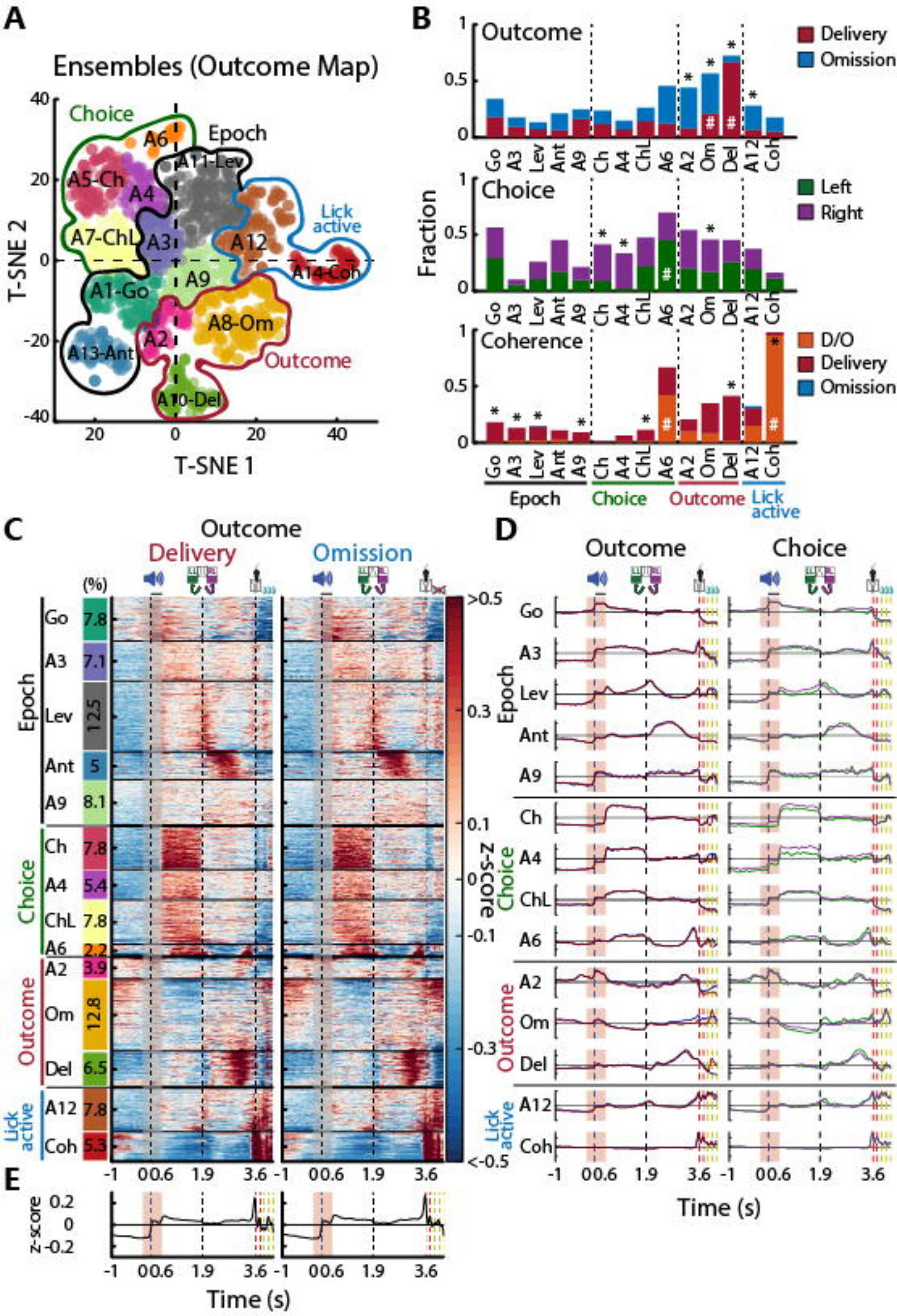

Right selective

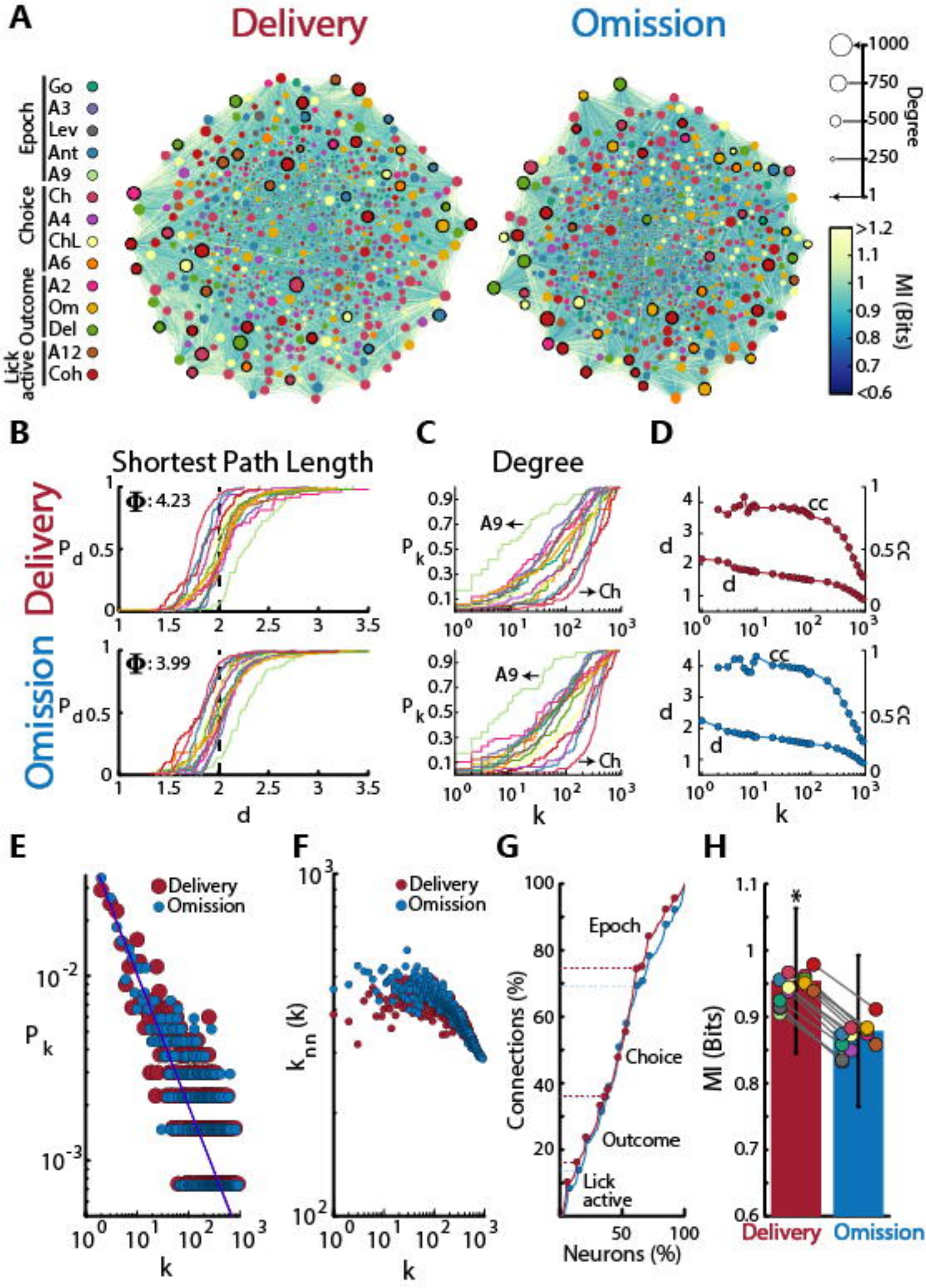
# Coherent cells

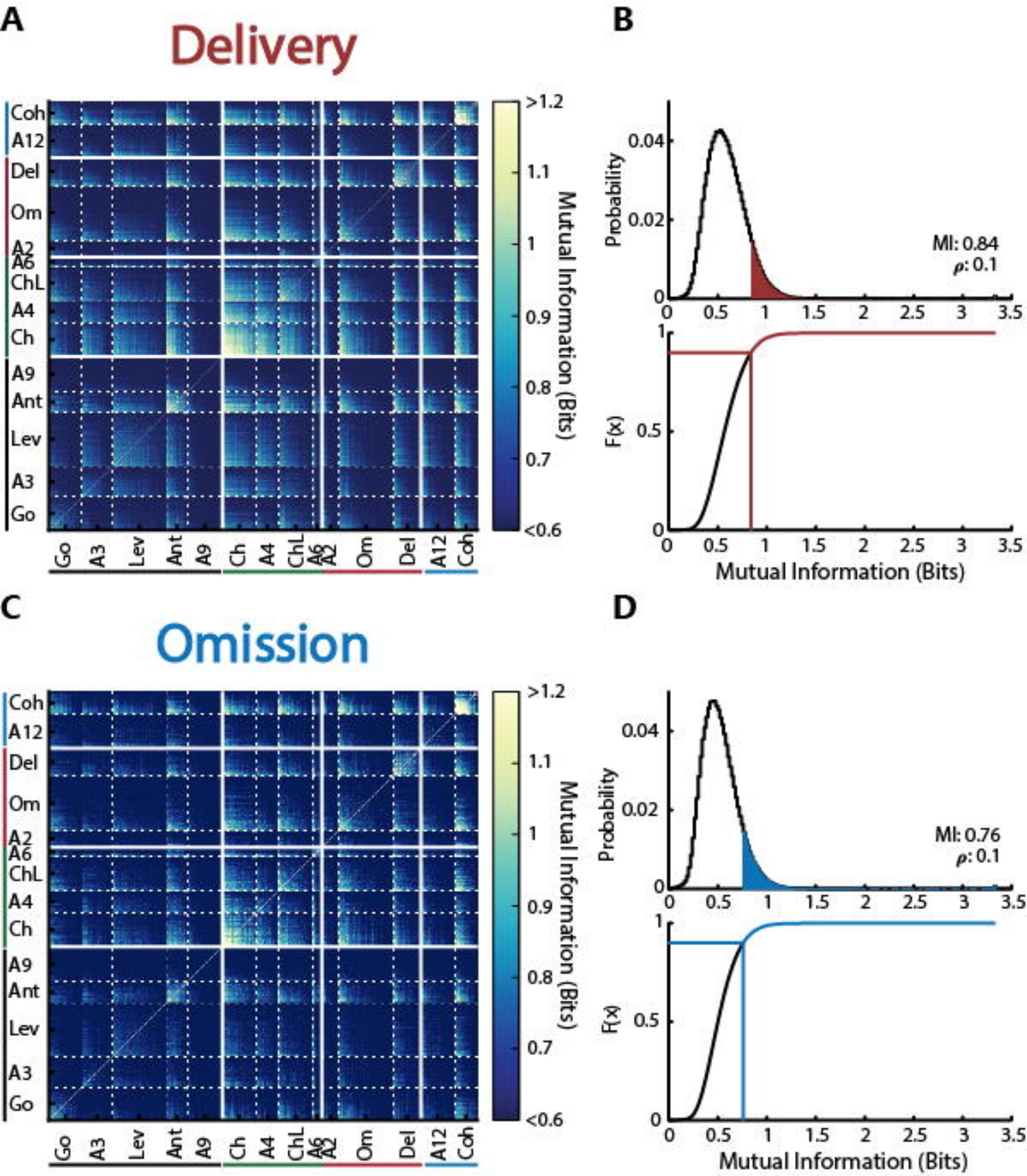
C

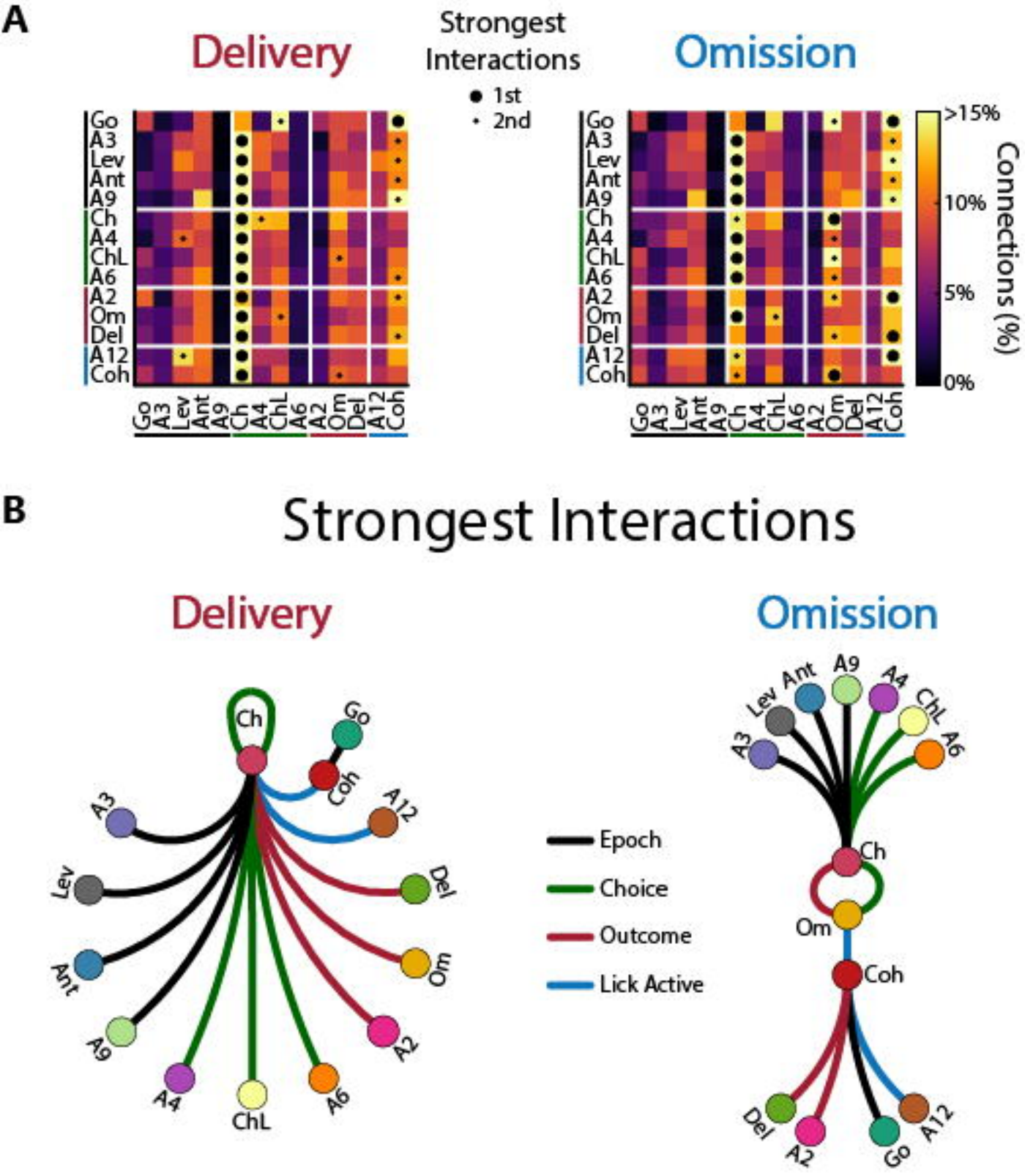


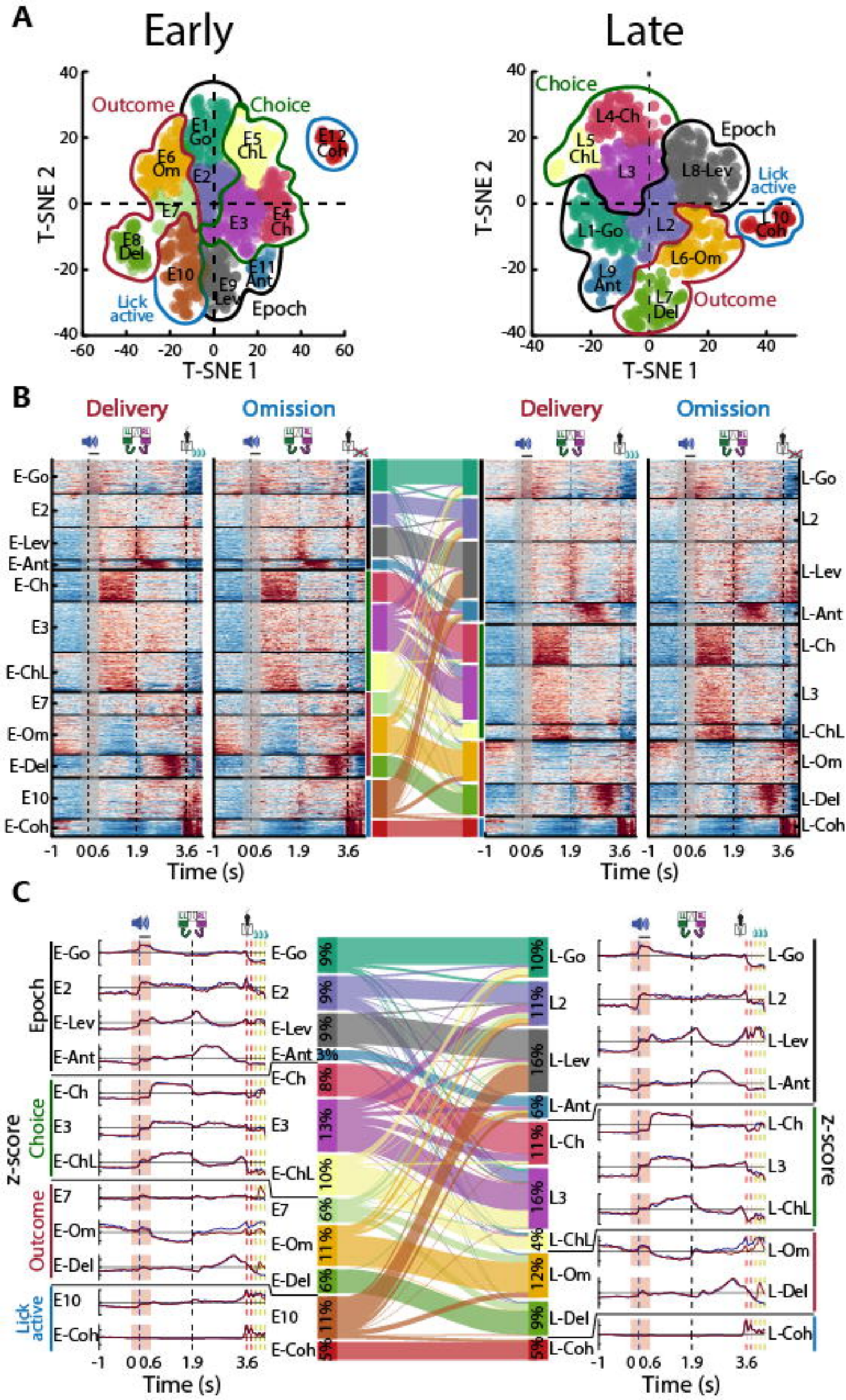

A
Delivery
Omission
Base Line
Go
Choice
Anticipation
Outcome
Neurons (Max Peak Activity)
Go BL
Choice
Anticipation
Outcome
Time (s)
z-score
B
Delivery
Omission
Go 8%
BL 4%
Outcome 20%
Choice 27%
Anticipation 41%
Go 7%
BL 2%
Outcome 17%
Choice 27%
Anticipation 46%
C
Delivery
Baseline 28.07 5.26 1.75 26.32 38.6
Go 1.74 48.7 18.26 25.22 6.09
Choice 0.25 3.69 70.52 15.72 9.83
Anticipation 0.64 2.57 9.79 79.13 7.87
Outcome 0.66 6.56 14.43 30.82 47.54
Baseline Go Choice Anticipation Outcome
Omission
% neurons mantained max peak activity in task epoch
D
PC 1 29.89%
PC 2 16.59%
PC 3 10.58%
PC 4 7.11%
Sucrose
Delivery
Omission
Time (s)
E
Sucrose
Delivery
Omission
Outcome
Dry Lick
Protrution
Anticipation
BL
Entry
Go
Lever
Exit
Choice
PC 1
PC 4
PC 3
F
G
Euclidean distance from zero
Delivery vs Omission
Time (s)

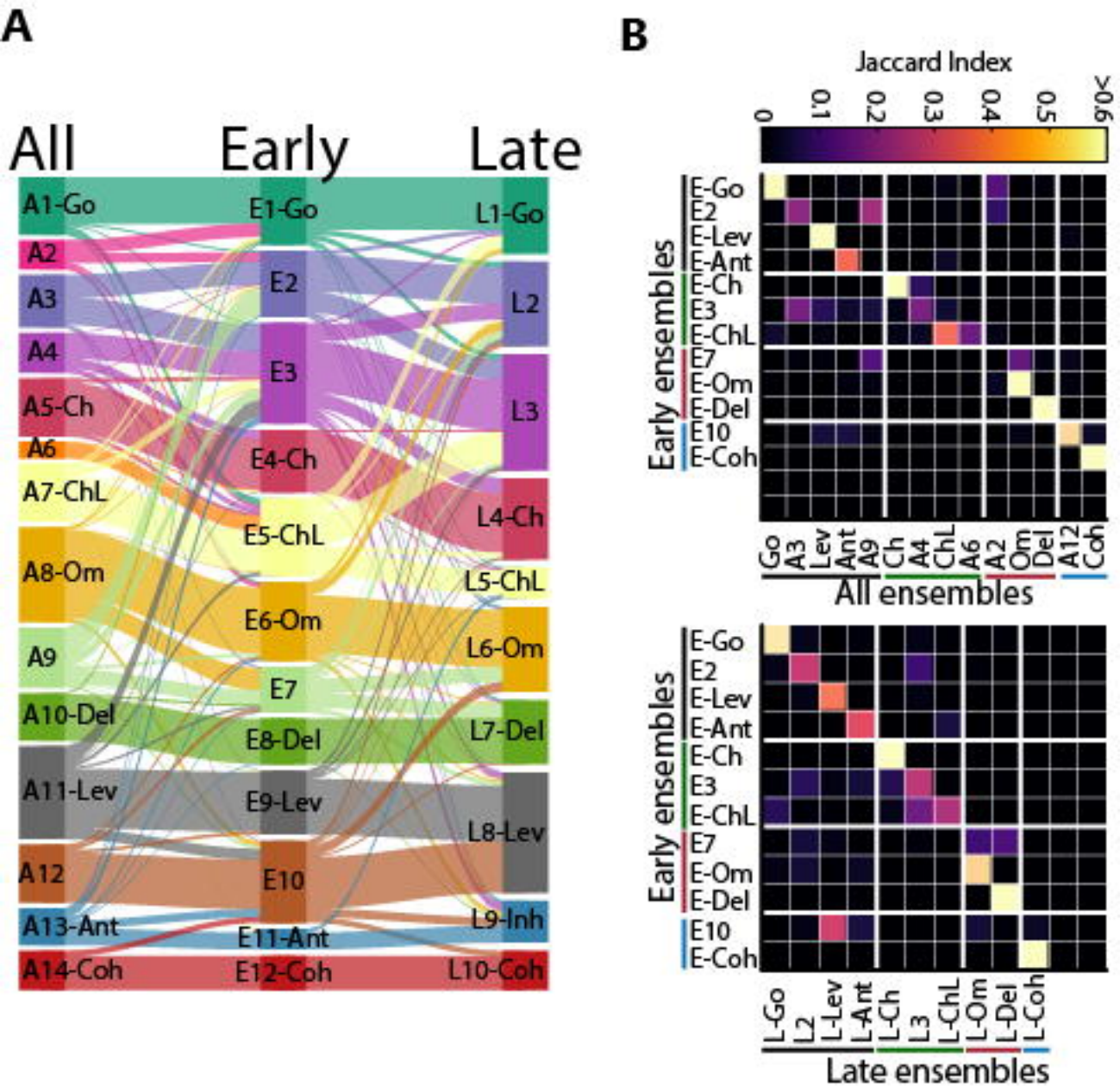

A
Ensembles (Outcome Map)
Choice
Epoch
Lick active
Outcome
A6
A11-Lev
A5-Ch
A4
A7-ChL
A3
A12
A14-Coh
A9
A1-Go
A8-Om
A2
A13-Ant
A10-Del
T-SNE 1
T-SNE 2
B
Outcome
Delivery
Omission
Choice
Left
Right
Coherence
D/O
Fraction
Go A3 Lev Ant A9 Ch A4 ChL A6 A2 Om Del A12 Coh
Epoch
Choice
Outcome
Lick active
C
Outcome
Delivery
Omission
(%)
Epoch
Go 7.8
A3 7.1
Lev 12.5
Ant 5
A9 8.1
Choice
Ch 7.8
A4 5.4
ChL 7.8
A6 2.2
Outcome
A2 3.9
Om 12.8
Del 6.5
Lick active
A12 7.8
Coh 5.3
>0.5
0.3
0.1
0
-0.1
-0.3
<-0.5
z-score
-1 00.6 1.9 3.6
D
Outcome
Choice
Time (s)
E
z-score
0.2
0
-0.2
Time (s)

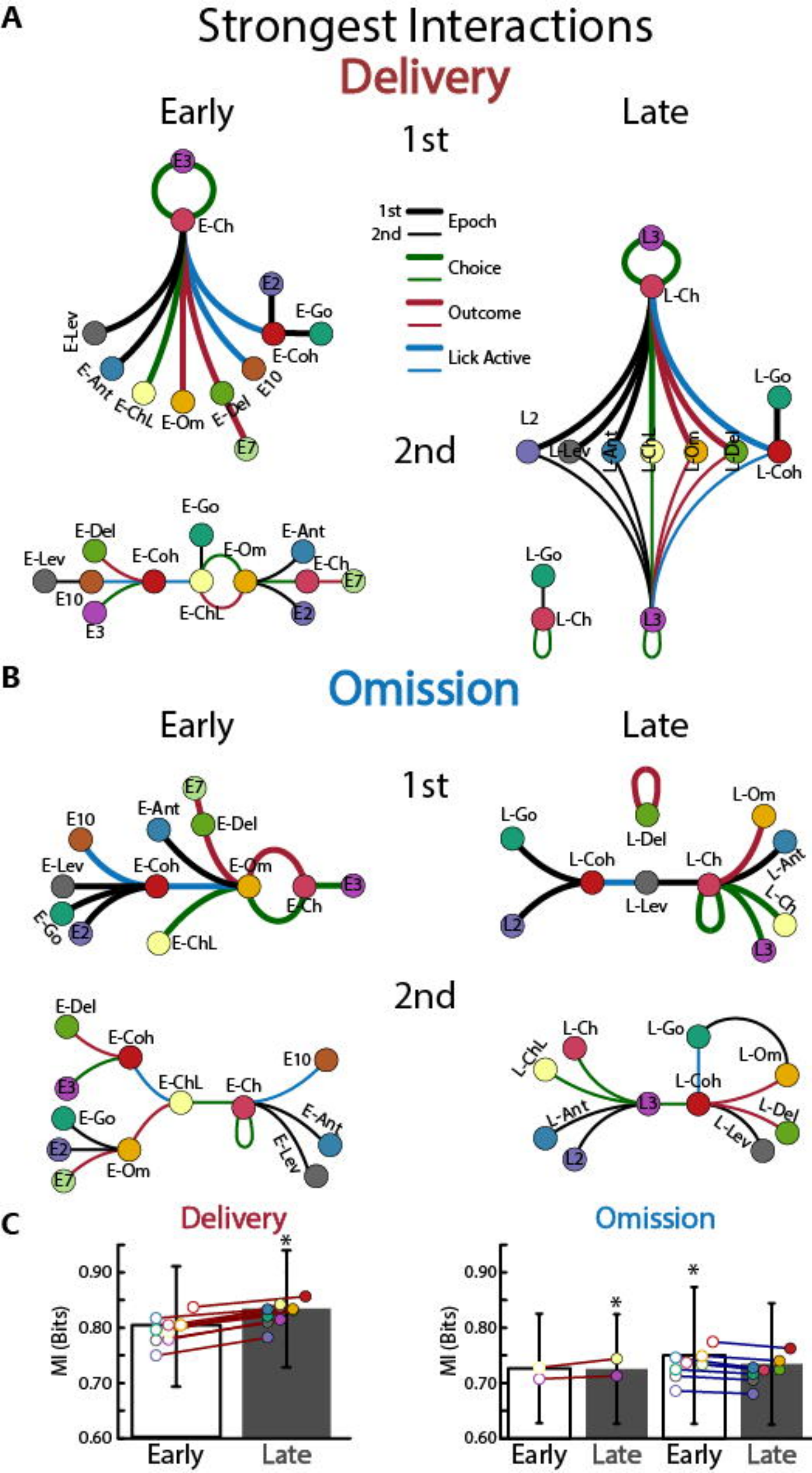

A
Delivery
Omission
Epoch
Go
A3
Lev
Ant
A9
Choice
Ch
A4
ChL
A6
Outcome
A2
Om
Del
Lick active
A12
Coh
1000
750
500
250
1
Degree
>1.2
1.1
1
0.9
0.8
0.7
<0.6
MI (Bits)
B
Shortest Path Length
Delivery
Omission
Φ: 4.23
Φ: 3.99
P_d
d
C
Degree
A9
Ch
P_k
k
D
cc
d
E
Delivery
Omission
P_k
k
F
k_nn (k)
k
G
Epoch
Choice
Outcome
Lick active
Connections (%)
Neurons (%)
H
MI (Bits)
Delivery
Omission

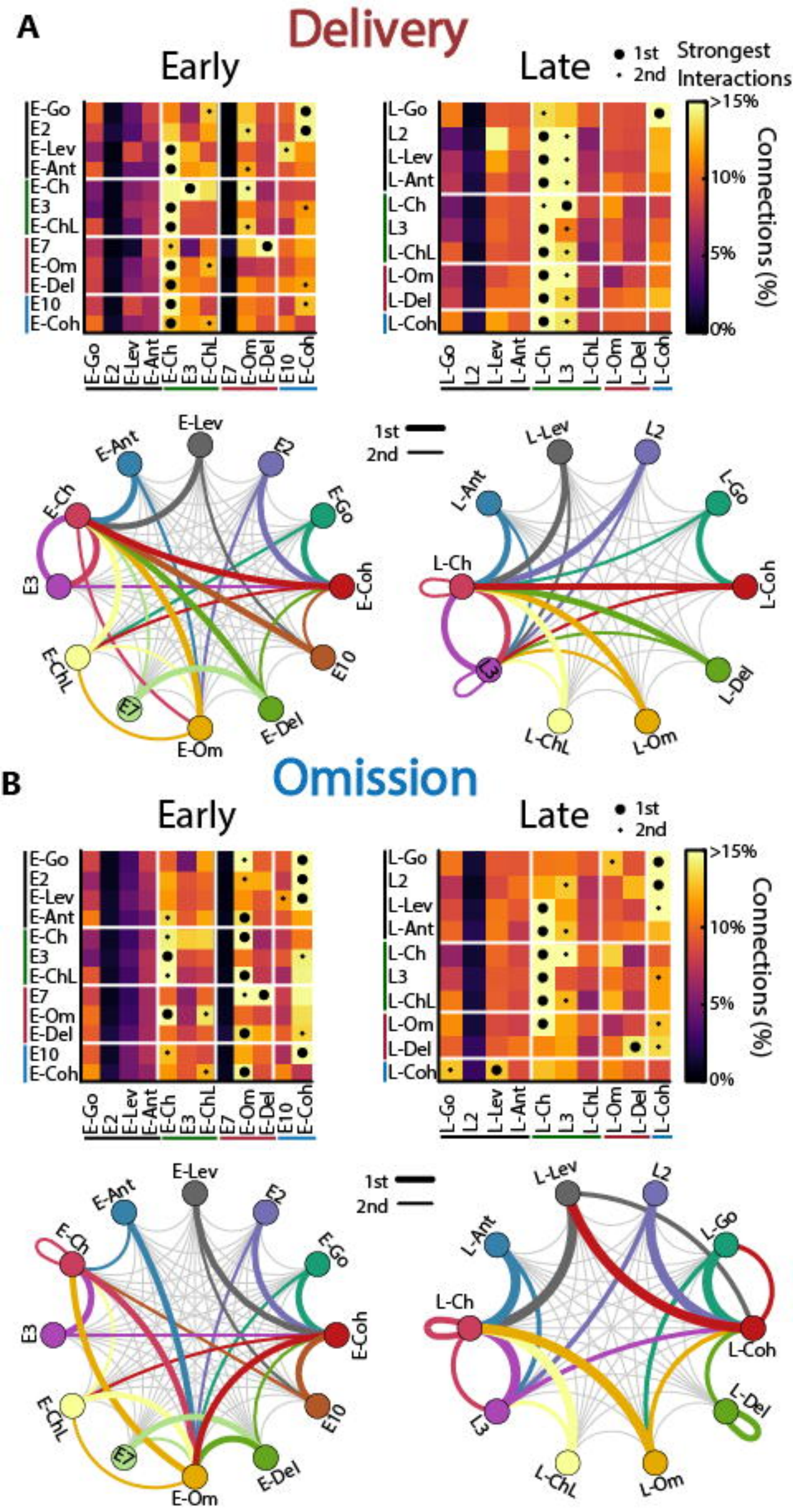

A
Delivery
B
Coh
A12
Del
Om
A2
A6
ChL
A4
Ch
A9
Ant
Lev
A3
Go
>1.2
1.1
1
0.9
0.8
0.7
<0.6
Mutual Information (Bits)
Probability
0.04
0.02
0
MI: 0.84
ρ: 0.1
0.5
1
1.5
2
2.5
3
3.5
F(x)
Mutual Information (Bits)
C
Omission
D
MI: 0.76
ρ: 0.1

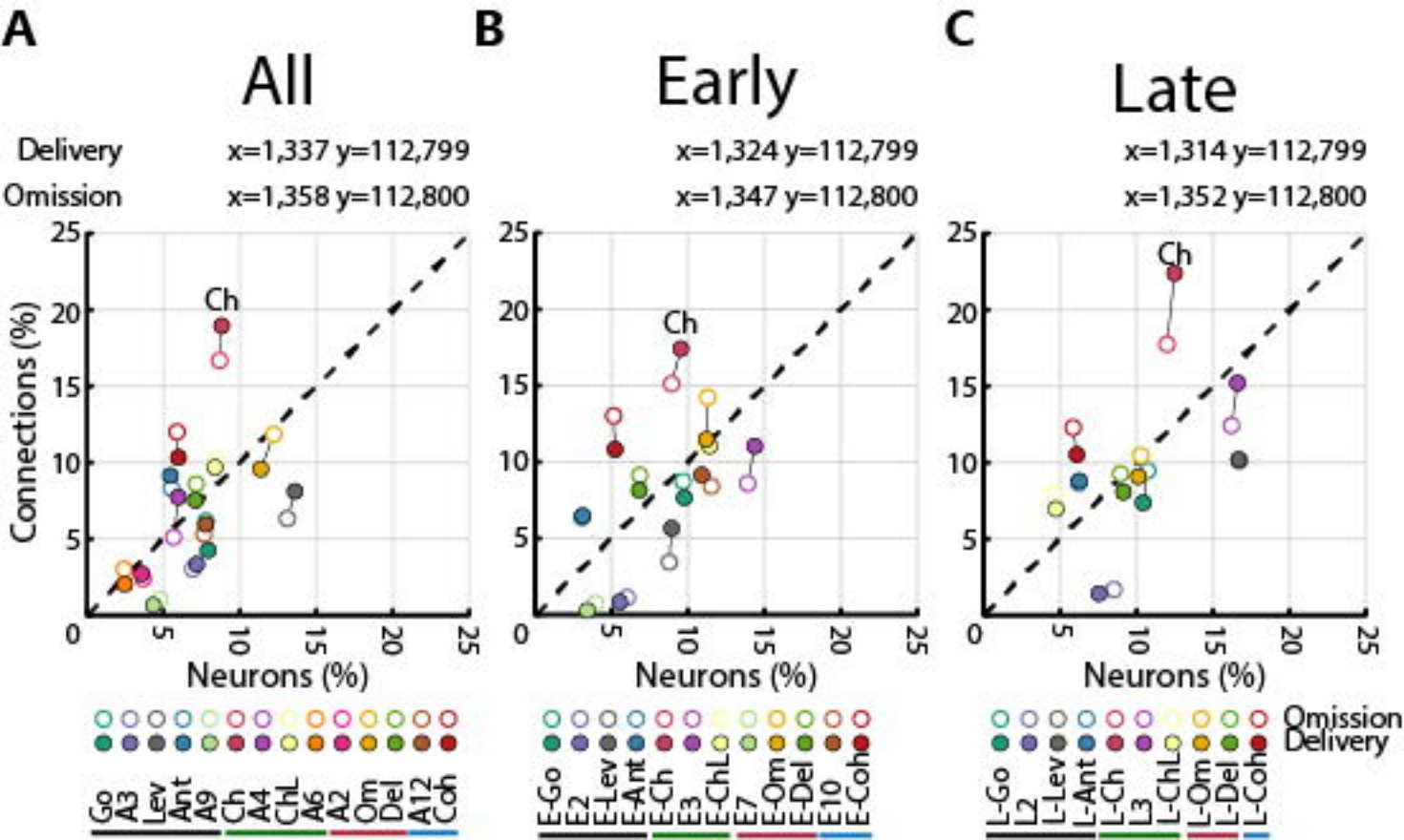

A
Delivery
Strongest Interactions
1st
2nd
Omission
Go
A3
Lev
Ant
A9
Ch
A4
ChL
A6
A2
Om
Del
A12
Coh
>15%
10%
5%
0%
Connections (%)
B
Strongest Interactions
Delivery
Omission
Epoch
Choice
Outcome
Lick Active

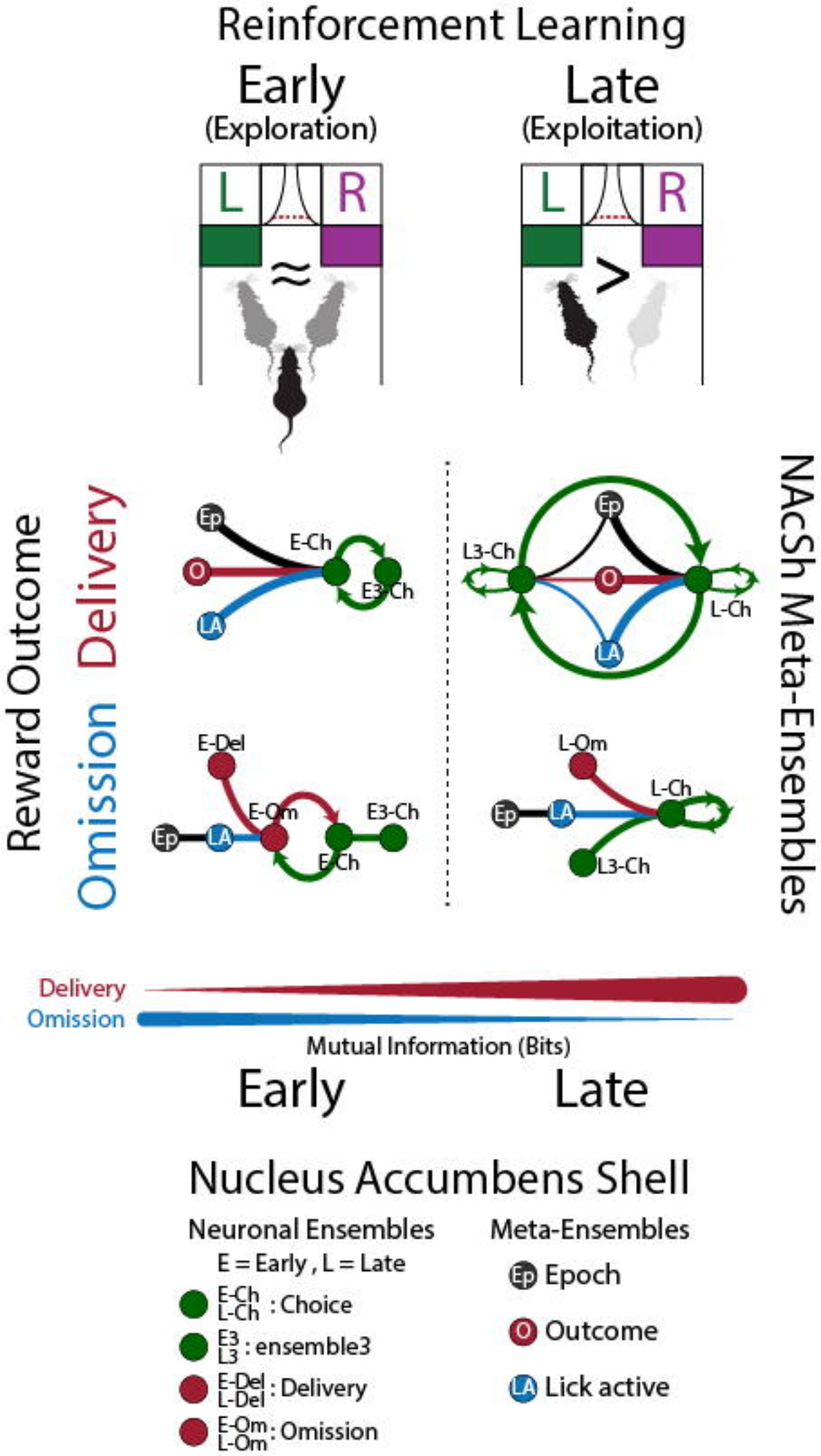

A
Early
Late
Outcome
Choice
Epoch
Lick active
E1 Go
E2
E3
E4 Ch
E5 ChL
E6 Om
E7
E8 Del
E9 Lev
E10
E11 Ant
E12 Coh
L1-Go
L2
L3
L4-Ch
L5 ChL
L6-Om
L7 Del
L8-Lev
L9 Ant
L10 Coh
T-SNE 1
T-SNE 2
B
Delivery
Omission
E-Go
E2
E-Lev
E-Ant
E-Ch
E3
E-ChL
E7
E-Om
E-Del
E10
E-Coh
L-Go
L2
L-Lev
L-Ant
L-Ch
L3
L-ChL
L-Om
L-Del
L-Coh
Time (s)
C
Epoch
Choice
Outcome
Lick active
z-score
9%
9%
9%
3%
8%
13%
10%
6%
11%
6%
11%
5%
10%
11%
16%
6%
11%
16%
4%
12%
9%
5%


A
All
Early
Late
A1-Go
A2
A3
A4
A5-Ch
A6
A7-ChL
A8-Om
A9
A10-Del
A11-Lev
A12
A13-Ant
A14-Coh
E1-Go
E2
E3
E4-Ch
E5-ChL
E6-Om
E7
E8-Del
E9-Lev
E10
E11-Ant
E12-Coh
L1-Go
L2
L3
L4-Ch
L5-ChL
L6-Om
L7-Del
L8-Lev
L9-Inh
L10-Coh
B
Jaccard Index
0
0.1
0.2
0.3
0.4
0.5
>0.6
Early ensembles
E-Go
E2
E-Lev
E-Ant
E-Ch
E3
E-ChL
E7
E-Om
E-Del
E10
E-Coh
Go
A3
Lev
Ant
A9
Ch
A4
ChL
A6
A2
Om
Del
A12
Coh
All ensembles
Early ensembles
E-Go
E2
E-Lev
E-Ant
E-Ch
E3
E-ChL
E7
E-Om
E-Del
E10
E-Coh
L-Go
L2
L-Lev
L-Ant
L-Ch
L3
L-ChL
L-Om
L-Del
L-Coh
Late ensembles


A
Strongest Interactions
Delivery
Early
Late
1st
2nd
1st
2nd
Epoch
Choice
Outcome
Lick Active
E3
E-Ch
E2
E-Go
E-Coh
E-Lev
E-Ant
E-ChL
E-Om
E-Del
E10
E7
L3
L-Ch
L-Go
L2
L-Lev
L-Ant
L-ChL
L-Om
L-Del
L-Coh
B
Omission
Early
Late
1st
2nd
C
Delivery
Omission
MI (Bits)
0.90
0.80
0.70
0.60
Early
Late


A
Delivery
Early
Late
1st
2nd
Strongest Interactions
>15%
10%
5%
0%
Connections (%)
E-Go
E2
E-Lev
E-Ant
E-Ch
E3
E-ChL
E7
E-Om
E-Del
E10
E-Coh
L-Go
L2
L-Lev
L-Ant
L-Ch
L3
L-ChL
L-Om
L-Del
L-Coh
B
Omission


A
All
Delivery x=1,337 y=112,799
Omission x=1,358 y=112,800
Ch
Connections (%)
Neurons (%)
B
Early
x=1,324 y=112,799
x=1,347 y=112,800
Ch
Neurons (%)
C
Late
x=1,314 y=112,799
x=1,352 y=112,800
Ch
Neurons (%)
Omission
Delivery
Go A3 Lev Ant A9 Ch A4 ChL A6 A2 Om Del A12 Coh
E-Go E2 E-Lev E-Ant E-Ch E3 E-ChL E7 E-Om E-Del E10 E-Coh
L-Go L2 L-Lev L-Ant L-Ch L3 L-ChL L-Om L-Del L-Coh


Reinforcement Learning
Early
(Exploration)
Late
(Exploitation)
L
R
≈
>
Reward Outcome
Delivery
Omission
Ep
E-Ch
O
E3-Ch
LA
L3-Ch
L-Ch
E-Del
E-Om
L-Om
NAcSh Meta-Ensembles
Delivery
Omission
Mutual Information (Bits)
Early
Late
Nucleus Accumbens Shell
Neuronal Ensembles
E = Early , L = Late
E-Ch L-Ch : Choice
E3 L3 : ensemble3
E-Del L-Del : Delivery
E-Om L-Om : Omission
Meta-Ensembles
Ep Epoch
O Outcome
LA Lick active